

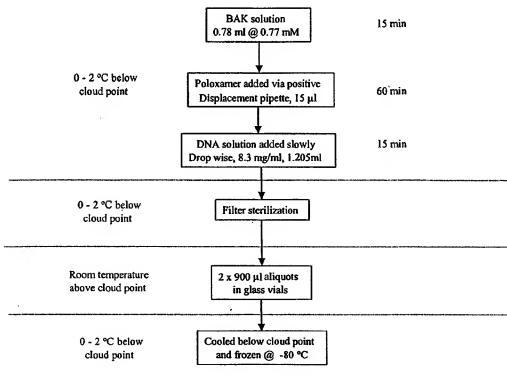
**EXHIBIT 2 OF DECLARATION UNDER
37 C.F.R § 1.131**



US 2007/0105193A1

(19) **United States**(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0105193 A1**
(43) **Pub. Date: May 10, 2007**(54) **SEVERE ACUTE RESPIRATORY
SYNDROME DNA VACCINE
COMPOSITIONS AND METHODS OF USE****Publication Classification**(75) **Inventors:** Adrian Vilalta, San Diego, CA (US);
Thomas G. Evans, San Diego, CA
(US); Melanie W. Quong, San Diego,
CA (US); Marston Manthorpe, San
Diego, CA (US)(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C12P 21/06 (2006.01)
(52) **U.S. CL.** 435/69.1; 435/6**Correspondence Address:**
STERNE, KESSLER, GOLDSTEIN & FOX
P.L.L.C.
1100 NEW YORK AVENUE, N.W.
WASHINGTON, DC 20005 (US)(73) **Assignee:** Vical Incorporated, San Diego, CA (US)(21) **Appl. No.:** 10/843,656(22) **Filed:** May 12, 2004**Related U.S. Application Data**(60) **Provisional application No. 60/482,505, filed on Jun.
26, 2003. Provisional application No. 60/470,820,
filed on May 16, 2003.**(57) **ABSTRACT**

The present invention is directed to raising a detectable immune response in a vertebrate by administering *in vivo*, into a tissue of the vertebrate, at least one polynucleotide comprising one or more regions of nucleic acid encoding a SARS-CoV protein or a fragment, a variant, or a derivative thereof. The present invention is further directed to raising a detectable immune response in a vertebrate by administering *in vivo*, into a tissue of the vertebrate, at least one SARS-CoV protein or a fragment, a variant, or derivative thereof. The SARS-CoV protein can be, for example, in purified form. The polynucleotide is incorporated into the cells of the vertebrate *in vivo*, and an immunologically effective amount of an immunogenic epitope of a SARS-CoV polypeptide, fragment, variant, or derivative thereof is produced *in vivo*. The SARS-CoV protein is also administered in an immunologically effective amount.



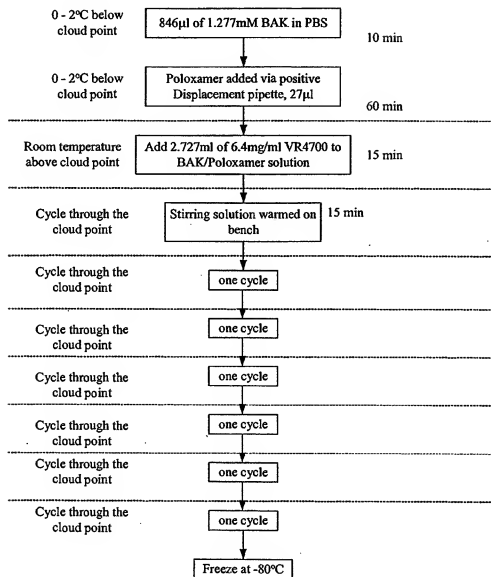


FIG. 1

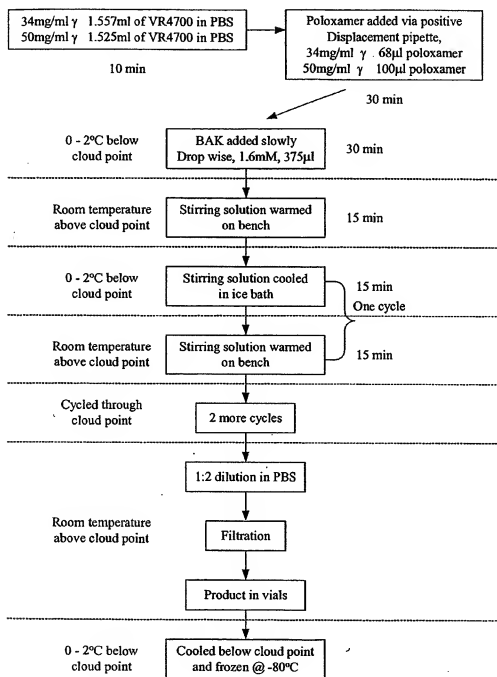


FIG. 2

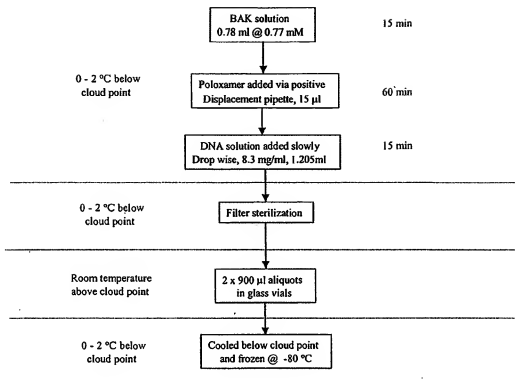


FIG. 3

SEVERE ACUTE RESPIRATORY SYNDROME DNA VACCINE COMPOSITIONS AND METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of the filing date of U.S. Provisional Application No. 60/470,820, filed May 16, 2003, and U.S. Provisional Application No. 60/482,505, filed Jun. 26, 2003, which are both incorporated herein by reference in their entirety.

BACKGROUND OF THE INVENTION

[0002] The present invention relates to a novel coronavirus (referred to herein as SARS-CoV) and to SARS-CoV vaccine compositions and methods of treating or preventing SARS-CoV infection and disease in mammals. SARS-CoV was discovered in March of 2003, in association with Severe Acute Respiratory Syndrome (SARS), a newly emerging infectious disease of global importance.

[0003] The recognition of SARS has led to activation of a global response network, with resultant travel restrictions, major quarantine, and closure of health care facilities. As of May 14, 2003, 7628 cases and 587 deaths from SARS have been reported from 29 countries. Initial reports of an atypical pneumonia began to surface in November of 2002 from the Guangdong province of China. This early outbreak reportedly involved 305 people, many of whom were healthcare workers. On Feb. 21, 2003, a healthcare worker from Guangdong traveled to Hong Kong, where his pre-existing cold symptoms escalated and he was hospitalized for acute respiratory distress. From Hong Kong, the illness spread rapidly throughout Southeast Asia and to Canada from this one index case. Seven individuals can be linked to the index case through a stay on the ninth floor of the hotel he occupied during his first night in Hong Kong. Infected persons from three hospitals in the Hong Kong metropolitan area are traceable to this index case as well. The primary mode of transmission has been either person-to-person contact or droplet transmission. Two notable exceptions to this are the hotel in Hong Kong, where direct human contact cannot be established for all those infected, and the Amoy Garden apartment buildings where more than 221 residents have been infected. In the outbreak at the Amoy Garden apartments, an unknown environmental factor is suspected of playing a role in transmission.

[0004] The incubation period ranges on average between two and seven days. Onset of symptoms begins with a high fever associated with chills and rigors. Additional symptoms at onset may include headache, malaise, myalgia, mild respiratory symptoms and more rarely common cold symptoms such as sore throat and runny nose. After this initial three to seven day period, additional lower respiratory symptoms appear including dry, non-productive cough and dyspnea. Initial chest x-rays reveal small, unilateral, patchy shadowings that progress quickly to bilateral, diffuse infiltrates. Preliminary Outbreak news: severe acute respiratory syndrome (SARS). *Wkly. Epidemiol. Rec.*, 2003: 81-88 (2003). The median duration of symptoms in a small epidemiologic study was 25.5 days. Tsang, K. W., et al. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* (2003). The severity of illness

can range widely from a mild illness to acute respiratory failure resulting in death. Patients with a significant comorbidity, such as diabetes, or who are older, are more likely to suffer from a severe form of the disease. Questions remain as to why some patients become infected, while others who have intimate contact with infected individuals are spared. It does appear that patients are very contagious at the onset of symptoms. Studies from hospitals in Hong Kong and Hanoi have shown attack rates >56% among healthcare workers caring for SARS patients. It is unclear at this time whether individuals are contagious during the incubation phase.

Important Features of Coronaviruses

[0005] Coronaviruses are large, enveloped, positive-stranded RNA viruses, and they are known to elicit coincident diseases in animals and humans. Mature human coronavirus (HCoV) virions are approximately 100 nm-diameter enveloped particles exposing prominent spike (S), hemagglutinin-esterase (HE) (in some types of coronaviruses), envelope (E) and membrane (M) glycoproteins. Each particle contains an approximately 30 kilobase (kB) RNA genome complexed with an approximately 60 kilodalton (kD) nucleoprotein (N). Fields, B. N. *VIROLOGY* New York: Lippincott, Williams & Wilkins, (Fields, B. N., ed. 2001). All of the above references are herein incorporated by reference in their entirety.

[0006] The S proteins of HCoV's have two large domains, the variable SI domain responsible for host cell binding, Breslin, J. J. et al. *J. Virol.* 77: 4435-8 (2003), and the S2 domain containing a heptad coiled-coiled structure reminiscent of those involved in fusion in HIV and influenza. Yoo, D. W. et al. *Virology* 183: 91-8 (1991). The HCoV-229E, group 1 S protein appears to bind to the human aminopeptidase N glycoprotein, Yeager, C. L., et al. *Nature* 357: 420-2 (1992); Bonavia, A. et al. *J. Virol.* 77: 2530-8 (2003), whereas the HCoV-OC43 strain (HCoV-OC43, group II) may bind via sialic acid moieties. Vlasak, R. et al. *Proc. Natl. Acad. Sci. USA* 85:4526-9 (1988). The genetic variability between strains of coronavirus has not been thoroughly evaluated, although only minor variability has been observed in the S protein in the small number of strains sequenced. Hays, J. P. and Myint, S. H. *J. Virol. Methods* 75: 179-93 (1998); Kunkel, F. and Herler, G. *Arch. Virol.* 141: 1123-31 (1996). Most coronaviruses are not only species specific, but also somewhat tissue tropic. This tropism is mostly related to changes in the S protein. Sanchez, C. M. et al. *J. Virol.* 73: 7607-18 (1999). Examples of such coronavirus tropism changes are the in vitro demonstration that tropism can be experimentally manipulated by genetically replacing a feline S protein with a mouse S protein, and the natural emergence of the porcine respiratory coronavirus (PRCoV) from the transmissible gastroenteritis virus of swine (TGEV) strain merely through a deletion of a region in the S protein. Haijema, B. J. et al. *J. Virol.* 77:4528-38 (2003); Page, K. W. et al. *J. Gen. Virol.* 72:579-87 (1991); Britton, P. et al. *Virus Res.* 21:181-98 (1991). All of the above references are herein incorporated by reference in their entirety.

[0007] The recently discovered novel coronavirus, SARS-CoV, appears to be a new member of the order Nidovirales. Concerted efforts by many laboratories worldwide has led to the rapid sequencing of various strains of SARS-CoV, including CUKH-Su10 (GenBank Accession No.

AY282752), TOR2 (GenBank Accession No. AY274119 and NC_004781), BJ01 (GenBank Accession No. AY278488), CUHK-W1 (GenBank Accession No. AY278554), Urbani (GenBank Accession No. AY278741) and HKU-39849 (GenBank Accession No. AY278491). The Urbani strain of SARS-CoV, sequenced by the Centers for Disease Control in Atlanta, Ga., is a 29,727-nucleotide, polyadenylated RNA with a genomic organization that is typical of coronaviruses: 5'-replicase, spike (S), envelope (E), membrane (M)-3'. Rota et al., *Science* 300:1394-1399 (2003), available May 1, 2003 at <http://www.sciencexpress.org> (hereinafter "Rota et al."). In addition, there are short untranslated regions at both termini, and open reading frames (ORFs) encoding non-structural proteins located between S and E, between M and N, or downstream of N. Rota et al. The hemagglutinin-esterase (HE) gene found in group 2 and some group 3 coronaviruses was not found in SARS-CoV. Rota et al. Sequencing of the Tor2 SARS-CoV strain by a collaboration of researchers in British Columbia, Canada, yielded a genomic sequence that differed from the Urbani SARS-CoV strain by eight nucleotide bases. Marra et al., *Science* 300:1399-1404 (2003), available May 1, 2003 at <http://www.sciencexpress.org> (hereinafter "Marra et al."). A comparison of the HKU-39849 and CUHK-W1 SARS-CoV strains also differed from the Urbani sequence by 10 or fewer nucleotide bases. Rota et al. All of the above references are herein incorporated by reference in their entirety.

[0008] Phylogenetic analyses indicate that, based on the genetic distance between SARS-CoV and other known coronaviruses in all of their genetic regions, no large region of the SARS-CoV genome was derived from other known viruses, and that SARS forms a distinct group within the genus *Coronavirus*. Rota et al.; Marra et al. The analyses also showed greater sequence conservation among enzymatic proteins of SARS-CoV than among the S, N, M, and E structural proteins; and, while there were regions of amino acid conservation within each protein as between SARS-CoV and other coronaviruses, the overall similarity was low. Rota et al. All of the above references are herein incorporated by reference in their entirety.

[0009] A virus, almost identical to the human SARS-CoV virus, has been isolated from rare Chinese masked palm civet cats. This virus is believed to be identical to human SARS-CoV except for a 29 nucleotide deletion in the region encoding the N protein of the virus. Walgate, R. "Human SARS virus not identical to civet virus" *The Scientist*, May 27, 2003, available at <http://www.biomedcentral.com/news/20030527/03/> (visited Jun. 13, 2003), incorporated herein by reference in its entirety.

Coronavirus Vaccine Candidates

[0010] Because SARS-CoV was so recently discovered, there are no vaccines against the virus. The approach to vaccine development can, however, be partially guided by the results of past studies in animals, of which three diseases have received the greatest attention. These are transmissible gastroenteritis virus (TGEV) in swine, feline infectious peritonitis virus (FIPV), and avian infectious bronchitis virus (IBV). Of note, none of the vaccines, most of which have been attenuated vaccines, have proven to be highly efficacious except for inactivated IBV. Enjuanes, L. et al., *Adv. Exp. Med. Biol.* 380: 197-211 (1995). The FIPV vaccine is a five attenuated virus that has provided minimal

efficacy in field trials, and the TGEV vaccine has also been problematic. Scott, F. W., *Adv. Vet. Med.* 41:347-58 (1999); Sestak, K. et al., *Vet. Immunol. Immunopathol.* 70:203-21 (1999). All of the above references are herein incorporated by reference in their entirety.

[0011] In the TGEV model, the major focus has been on neutralizing antibody directed at the S glycoprotein. Sestak, K. et al., *Vet. Immunol. Immunopathol.* 70: 203-21 (1999); Tuboly, T. et al. *Vaccine* 18: 2023-8 (2000); Shoup, D. I. et al. *Am. J. Vet. Res.* 58: 242-50 (1997). Protection has also been associated with antibodies in IBV and bovine coronavirus. Mondal, S. P. et al. *Avian. Dis.* 45:1054-9 (2001); Yoo, D. W. et al. *Virology* 180: 395-9 (1991). In fact, in most of the animal models, control of coronavirus infection can be due to antibodies reactive to the N-terminal region of the S protein. Gallagher, T. M. and Buchmeier, M. J. *Virology* 279: 371-4 (2001); Tuboly, T. et al. *Arch. Virol.* 137: 55-67 (1994). In one study of respiratory bovine coronavirus, antibody appearance to the S and N proteins was correlated with recovery. Lin, X. Q. et al. *Arch. Virol.* 145: 2335-49 (2000); Passive transfer studies have also been successful and demonstrated the value of humoral immune responses. Enjuanes, L. et al., *Adv. Exp. Med. Biol.* 380: 197-211 (1995); Spaan, W. J. *Adv. Exp. Med. Biol.* 276: 201-3 (1999). All of the above references are herein incorporated by reference in their entirety.

[0012] Cell-mediated immune responses have been most clearly detected in coronaviruses against the S, M and N proteins. Spencer, J. S. et al., *Adv. Exp. Med. Biol.* 380: 121-9 (1995); Collisnon, E. W. et al. *Dev. Comp. Immunol.* 24: 187-200 (2000); Stohlman, S. A. et al. *Virology* 189: 217-24 (1992). In one study, the use of a DNA vaccine encoding the carboxyl terminus of the N gene of IBV, which induced cytotoxic T cell (CTL) activity, was able to decrease virus titers by 7 logs in target organs. Seo, S. H. et al. *J. Virol.* 71: 7889-94 (1997). Some protection was also noted in a DNA vaccine encoding the N protein in the Mouse Hepatitis Virus (MHV) model. Hayashi, M. et al. *Adv. Exp. Med. Biol.* 440:693-9 (1998). There is also some evidence that CTL may be involved in the control of MHV, and prevent the development of persistent infection and neuropathology. Pewe, L. and Perlman, S. *Virology* 255: 106-16 (1999); Pewe, L. et al. *J. Virol.* 71: 7640-7 (1997). All of the above references are herein incorporated by reference in their entirety.

[0013] A large number of coronavirus challenge studies have been conducted in humans by Tyrrell and colleagues, in which the subjects were inoculated intranasally and followed. Callow, K. A. et al. *Epidemiol. Infect.* 105: 435-46 (1990); Bende, M. et al. *Acta Otolaryngol.* 107: 262-9 (1989). Such challenge studies will clearly be impossible for the much more serious SARS-CoV virus. The presence of antibodies to the challenge strain did not prevent infection or disease, even in the face of rising neutralizing antibody titers. However, a second infection with similar strains led to decreased symptoms, revealing persistence of immunity against homologous challenge. Reed, S. E. *J. Med. Virol.* 13: 179-92 (1984). Also, the 2-4 year cyclical nature of the disease points to some persistence of immune response over time. Reed, S. E. *J. Med. Virol.* 13: 179-92 (1984); Hendley, J. O. et al. *Am. Rev. Respir. Dis.* 105: 805-11 (1972); Evans, A. S. and Kaslow, R. A. *VIRAL INFECTIONS OF HUMANS*, 4th ed. New York and London: Plenum Medical

Book Company, (Evans, A. S. and Kaslow, R. A., eds., 1997). All of the above references are herein incorporated by reference in their entireties.

[0014] Heterologous "prime boost" strategies have been effective for enhancing immune responses and protection against numerous pathogens. Schneider et al., *Immunol. Rev.* 170:29-38 (1999); Robinson, H. L., *Nat. Rev. Immunol.* 2:239-50 (2002); Gonzalo, R. M. et al., *Vaccine* 20:1226-31 (2002); Tanghe, A., *Infect. Immun.* 69: 3041-7 (2001). Providing antigen in different forms in the prime and the boost injections appears to maximize the immune response to the antigen. DNA vaccine priming followed by boosting with protein in adjuvant or by viral vector delivery of DNA encoding antigen appears to be the most effective way of improving antigen specific antibody and CD4+ T-cell responses or CD8+ T-cell responses respectively. Shiver J. W. et al., *Nature* 415: 331-5 (2002); Gilbert, S. C. et al., *Vaccine* 20:1039-45 (2002); Billaut-Mulot, O. et al., *Vaccine* 19:95-102 (2000); Sin, J. I. et al., *DNA Cell Biol.* 18:771-9 (1999). Recent data from monkey vaccination studies suggests that adding CRL1005 poloxamer to DNA encoding the HIV gag antigen enhances T-cell responses when monkeys are vaccinated with an HIV gag DNA prime followed by a boost with an adenoviral vector expressing HIV gag (Ad5-gag). The cellular immune responses for a DNA/poloxamer prime followed by an Ad5-gag boost were greater than the responses induced with a DNA (without poloxamer) prime followed by Ad5-gag boost or for Ad5-gag only. Shiver, J. W. et al., *Nature* 415:331-5 (2002). U.S. Patent Appl. Publication No. US 2002/0165172 A1 describes simultaneous administration of a vector construct encoding an immunogenic portion of an antigen and a protein comprising the said immunogenic portion of an antigen such that an immune response is generated. The document is limited to hepatitis B antigens and HIV antigens. Moreover, U.S. Pat. No. 6,500,432 is directed to methods of enhancing an immune response of nucleic acid vaccination by simultaneous administration of a polynucleotide and polypeptide of interest. According to the patent, simultaneous administration means administration of the polynucleotide and the polypeptide during the same immune response, preferably within 0-10 or 3-7 days of each other. The antigens contemplated by the patent include, among others, those of Hepatitis (all forms), HSV, HIV, CMV, EBV, RSV, VZV, HPV, polio, influenza, parasites (e.g., from the genus *Plasmodium*), pathogenic bacteria (including but not limited to *M. tuberculosis*, *M. leprae*, *Chlamydia*, *Shigella*, *B. burgdorferi*, enterotoxigenic *E. coli*, *S. typhosa*, *H. pylori*, *V. cholerae*, *B. pertussis*, etc.). All of the above references are herein incorporated by reference in their entireties.

SUMMARY OF THE INVENTION

[0015] The present invention is directed to compositions and methods for raising a detectable immune response in a vertebrate against the infectious agent transmitting Severe Acute Respiratory Syndrome (SARS), by administering in vivo, into a tissue of a vertebrate, at least one polynucleotide comprising one or more nucleic acid fragments, wherein each nucleic acid fragment is a fragment of a coding region operably encoding a polypeptide, or a fragment, variant, or derivative thereof, or a fragment of a codon-optimized coding region operably encoding a polypeptide, or a fragment, variant, or derivative thereof, from a coronavirus which causes SARS (SARS-CoV). The present invention is

also directed to administering in vivo, into a tissue of the vertebrate the above-described polynucleotide and at least one isolated SARS-CoV polypeptide, or a fragment, variant, or derivative thereof. The isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof can be, for example, a recombinant protein, a purified subunit protein, a protein expressed and carried by a heterologous live or inactivated or attenuated viral vector expressing the protein. According to either method, the polynucleotide is incorporated into the cells of the vertebrate in vivo, and an amount of the SARS-CoV protein, or fragment or variant encoded by the polynucleotide sufficient to raise a detectable immune response is produced in vivo. The isolated protein or fragment, variant, or derivative thereof is also administered in an amount sufficient to raise a detectable immune response. The polynucleotide may be administered to the vertebrate either prior to, at the same time (simultaneously), or subsequent to the administration of the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof.

[0016] Also within the scope of the present invention are combinations of SARS-CoV polypeptides and polynucleotides that encode SARS-CoV polypeptides that assemble into virus-like particles (VLP). One such combination is, but is not limited to a combination of SARS-CoV S, M, and E polypeptides or fragments, variants, or derivatives thereof, and polynucleotides encoding SARS-CoV S, M, and E polypeptides or fragments, variants, or derivatives thereof.

[0017] In a specific embodiment, the invention provides polynucleotide (e.g., DNA) vaccines in which the single formulation comprises a SARS-CoV polypeptide-encoding polynucleotide vaccine as described herein. An alternative embodiment of the invention provides for a multivalent formulation comprising several (e.g., two, three, four, or more) SARS-CoV polypeptide-encoding polynucleotides, as described herein, within a single vaccine composition. The SARS-CoV polypeptide-encoding polynucleotides, fragments or variants thereof may be contained within a single expression vector (e.g., plasmid or viral vector) or may be contained within multiple expression vectors.

[0018] In a specific embodiment, the invention provides combinatorial polynucleotide (e.g., DNA) vaccines which combine both a polynucleotide vaccine and polypeptide (e.g., either a recombinant protein, a purified subunit protein, a viral vector expressing an isolated SARS-CoV polypeptide) vaccine in a single formulation. The single formulation comprises a SARS-CoV polypeptide-encoding polynucleotide vaccine as described herein, and optionally, an effective amount of a desired isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof. The polypeptide may exist in any form, for example, a recombinant protein, a purified subunit protein, or a viral vector expressing an isolated SARS-CoV polypeptide. The SARS-CoV polypeptide or fragment, variant, or derivative thereof encoded by the polynucleotide vaccine may be identical to the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof. Alternatively, the SARS-CoV polypeptide or fragment, variant, or derivative thereof encoded by the polynucleotide may be different from the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof.

[0019] The present invention further provides a method for generating, enhancing, or modulating a protective and/or therapeutic immune response to SARS-CoV in a vertebrate,

comprising administering to a vertebrate in need of therapeutic and/or preventative immunity one or more of the compositions described herein.

[0020] The invention also provides for antibodies specifically reactive with SARS Co-V polypeptides which have been produced from an immune response elicited by the administration, to a vertebrate, of polynucleotide and polypeptides of the present invention.

[0021] In one embodiment, purified monoclonal antibodies or polyclonal antibodies containing the variable heavy and light sequences are used as therapeutic and prophylactic agents to treat or prevent SARS-CoV infection by passive antibody therapy. In general, this will comprise administering a therapeutically or prophylactically effective amount of the monoclonal antibodies to a susceptible vertebrate or one exhibiting SARS Co-V infection.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0022] FIG. 1 shows the protocol for the preparation of a formulation comprising 0.3 mM BAK, 7.5 mg/ml CRL 1005, and 5 mg/ml of DNA in a final volume of 3.6 ml, through the use of thermal cycling.

[0023] FIG. 2 shows the protocol for the preparation of a formulation comprising 0.3 mM BAK, 34 mg/ml or 50 mg/ml CRL 1005, and 2.5 mg/ml DNA in a final volume of 4.0 ml, through the use of thermal cycling.

[0024] FIG. 3 shows the protocol for the simplified preparation (without thermal cycling) of a formulation comprising 0.3 mM BAK, 7.5 mg/ml CRL 1005, and 5 mg/ml DNA.

DETAILED DESCRIPTION OF THE INVENTION

[0025] The present invention is directed to compositions and methods for raising a detectable immune response in a vertebrate against the infectious agent transmitting Severe Acute Respiratory Syndrome (SARS), by administering in vivo, into a tissue of a vertebrate, at least one polynucleotide comprising one or more nucleic acid fragments, wherein each nucleic acid fragment is a fragment of a coding region operably encoding a polypeptide, or a fragment, variant, or derivative thereof, or a fragment of a codon-optimized coding region operably encoding a polypeptide, or a fragment, variant, or derivative thereof, from a coronavirus which causes SARS (SARS-CoV). The present invention is also directed to administering in vivo, into a tissue of the vertebrate the above-described polynucleotide and at least one isolated SARS-CoV polypeptide, or a fragment, variant, or derivative thereof. The isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof can be, for example, a recombinant protein, a purified subunit protein, a protein expressed and carried by a heterologous live or inactivated or attenuated viral vector expressing the protein. According to either method, the polynucleotide is incorporated into the cells of the vertebrate in vivo, and an amount of the SARS-CoV protein, or fragment or variant encoded by the polynucleotide sufficient to raise a detectable immune response is produced in vivo. The isolated protein or fragment, variant, or derivative thereof is also administered in an amount sufficient to raise a detectable immune response. The polynucleotide may be administered to the vertebrate either

prior to, at the same time (simultaneously), or subsequent to the administration of the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof.

[0026] In certain embodiments, the present invention provides for methods for raising a detectable immune response to polypeptides from a SARS-CoV virus, comprising administering to a vertebrate a polynucleotide which operably encodes a SARS-CoV polypeptide, wherein said polynucleotide is administered in an amount sufficient to elicit a detectable immune response to the encoded polypeptide.

[0027] The nucleotide and amino acid sequences of several SARS-CoV polypeptides have recently been determined. Several strains of human SARS-CoV (hSARS-CoV) have been sequenced. Sequences available on GenBank include the complete genomic sequences for SARS coronavirus strains CUKH-Su10, TOR2, BJ01, CUHK-W1, Urbani, and HKU-39849. SARS-CoV polypeptides from any of these strains are within the scope of the invention. Non-limiting examples of SARS-CoV polypeptides within the scope of the invention include the Spike (S), Nucleocapsid (N), Envelope (E), and Membrane glycoprotein (M) polypeptides, fragments, derivatives, (e.g., a TPA-S fusion), and variants thereof. As shown in Table 1 below, adapted from Rota et al., the various SARS-CoV strains that have been sequenced differ in various nucleotide base positions, some of which, as shown in Table 2 below, adapted from Marra et al., may result in a different amino acid residue. Thus, also within the scope of the invention are polypeptides that have different amino acids at those positions. The SARS-CoV polypeptide examples described below are from the Urbani strain of SARS-CoV, and are not meant to be limiting in terms of the scope of the invention.

TABLE 1

Comparison of Genomic Sequences of SARS-CoV Strains

Nucleotide Position ^a	Consensus	HKU-39849	CUHK-W1	Urbani	TOR2
2,601	T	C	*	*	*
7,746	G	*	T	*	*
7,919	C	*	*	T	*
7,930	G	A	*	*	*
8,387	G	C	*	*	*
8,417	G	C	*	*	*
9,404	T	*	C	*	*
9,479	T	*	C	*	*
13,404	G	A	*	*	*
13,495	T	G	*	*	*
16,622	C	*	*	T	*
17,564	T	*	G	*	*
17,846	C	*	T	*	*
18,065	G	A	*	*	*
19,064	R	A	G	G	A
21,721	G	*	A	*	*
22,222	T	*	C	*	*
23,220	T	*	*	*	G
24,872	T	*	*	C	*
25,298	G	*	*	*	A
25,569	T	A	*	*	*
26,600	C	T	*	*	*
26,857	T	*	*	C	*
27,827	T	*	C	*	*

[0028]

TABLE 2

Comparison of Tor2 and Urbani Strains of SARS-CoV and Corresponding Amino Acid Substitutions					
Nucleotide Position	Tor2 Base	Corresponding Amino Acid	Urbani Base	Corresponding Amino Acid	Protein
7,919	C	A	T	V	Rep1A
16,622	C	A	T	A	Rep1B
19,064	A	E	G	E	Rep1B
19,183	T	V	C	A	Rep1B
23,220	G	A	T	S*	Spike (S)
24,872	T	L	C	L	Spike (S)
25,298	A	R	G	G*	ORF 3
26,857	T	S	C	P*	M

*Non-conservative Amino Acid Substitution

[0029] From about nucleotide 21492 to about 25259 of the Urbani strain of the SARS-CoV genome encode the Spike (S) protein. (Bellini et al. SARS Coronavirus Urbani, complete genome. GenBank Accession No. AY278741.) The complete S protein is about 1255 amino acids in length (139.12 kDa) and is predicted, by analogy to other coronaviruses, to be a surface projection glycoprotein precursor. The S protein has several important biologic functions. Monoclonal antibodies against S can neutralize virus infectivity, consistent with the observation that S protein binds to cellular receptors. The S glycoprotein has several important biologic functions. Monoclonal antibodies against S can neutralize virus infectivity, consistent with the observation that S protein binds to cellular receptors. The S protein is encoded by the following polynucleotide sequence in the Urbani strain and is referred to herein as SEQ ID NO:22.

ATGTTATTTCCTTATTATTCCTACTCTACTAGTGTGTAACCTGGA
CGGGTCACCACTTTTGATGATGTTCAAGCTCCTTAATACACTCAACATA
CTCATCTATGAGGCGGCTTACTATCTCGATGAATTTTATGATCAGAC
ACTCTTATTAACTCAGAGTTAATTTCTCCATTTTATCTTAATGTATC
AGGGTTCTACTATTAACTACATGCTTTGGCAACCTTGATACCTTTTA
AGATGATGATTTATTTGCTGOCACAGAGAATCAATGTTGTCGCTGGT
TGGTTTGTGCTTACACAGAACAGCTGACAGCTGGTGAATATAT
TAACTACTACTAACTGTTGATACAGACATGTAACCTTGAATGTGTG
ACACCCCTTCTTCTGCTTTTCAACCCATGGGTACAGACACACTACT
ATGATATTGATATGATTAATTTGACCTTTGAGTACATATCTGATGC
CTTCTTCGCTGATGTTTCAAGAAATCAGTAATTTTAAACACTTACGAG
AGTTTGTGTTTAAATAAAGATGGGTTTCTCTATGTTTAAAGGGCTAT
CAACCTATAGATGTAGTCTGATCTACCTCTGGTTTAACTACTTTGAA
ACCTATTTTAAGTGGCTCTGCTATTAACATACAAATTTTAAAGGCA
TTCTTACAGCTTTTCACTGCTCAAGACATTTGGGCGACCTCAGCTGCA
GCTTATTTTGTGGCTATTAAAGCCACTACATTTATGCTCAAGTATGA
TGAAATGTGACAACTACAGATGCTGTGATGTTCTCAAACTCACTTG

-continued

CTGAACCAAAATGCTCTGTAAGAGCTTGAAGATGACAAAGAAATTAC
CAGAGCTCTAAATTCAGGGTGTTCCTCAGGAGATGTTGTGAGATTGCC
TAAATATACAACTTGTGTCTTTTGAGAGAGTTTATGACTACTAATAT
TCCTCTGTCTATGATGAGGAGAGAAATTTCTAATGTGTGCT
GATTACTCTGTCTCTACAACTCAACATTTTTTCAACCTTAAAGTCTA
TGGCTTTCTGCCACTAAGTGAATGATCTTCTCTCAATGTCTATG
CAGATCTTTTGTAGTCAAGGGAGATGATGAAGCAAAATGCGCCAGGA
CAAACTGGTGTATGCTGATATTAATTAATTAATGCGAGATGATTCAT
GGGTGTGTCTGCTGCTGAATACATGAGAACATGATGCTACTTCAACTG
GTAATTAATTAATTAATGATGATCTTACAGATGCAAGCTTAGCGCC
TTTGAGAGAGACATATTAATGTGCTTTCTCCCTGATGGCAAGCTTG
CACCCACCTGCTCTTAATGTGTTATGCGCATTAATGATTATGGTTTT
ACACCACTACTGCGATGCTACCACTCTACAGAGTTAGTACTTCTT
TTGAACTTTTAAATGCAAGCGCCAGGTTTGTGAGCAAAATATCCAC
TGACTTATTAGAACCAAGTGTCTCAATTTTAAATTTAAGGACTACTGG
TACTGGTGTGTTAATCTCTCTCAAGAGATTTCAACCATTTCAACAT
TTGGCGGTGATGTTCTGATTTCACTGATTCGCTGAGATCTCAAAACA
CTGAAATATTAGACTTTACCTGCTCTTCTTTGGGGTGTAGTGTAAAT
TACACTGGAACAAATGCTTCATGTAAGTGTGCTGTCTATATCAGATG
TTAATCTGACTGATGTTCTACAGCAATTCATGCACTCACTCAGCACA
GCTTGGCGCATATATTTACTTGAACCAATGATTCAGACTCAAGCGC
CTGCTTTATAGGAGCTGAGCATGTGCACTCTCTTATGATGCGACATTC
CTATGAGAGCTGGCAATTTGTGCTAGTATCCATAGAGTTTCTTATACGT
AGTATACGCAAAATATTTTGTGCTTATACTATCTTCTTATGCTGCTGA
TAGTCAATGCTTACTCTAATAACACATTTCTATACCTACTAATCTTT
CAATGACATTAACACAGAGTAATGCTGCTTCTATGCTTAAACCTCC
GTAGATTGTAATATGATCTCTGCGAGATTTCTACTGAATGTGCTAAT
GCTTCTCAATATGTAGCTTTTGCACACAACTAAAGTGACACTCTCAG
GTATGTGCTGCAACAGGATGCAACACAGTGAAGTGTGCTGCTAAGTC
AAACAAATGTACAAACCCCACTTTGAAATATTTTGGTGGTTTAAATTT
CTGCAAAATTAATCTGACCTCTTAAAGCCAACTAAGAGGCTCTTTATG
AGGACTGCTCTTAAATAGGTGACTGCTGCTGATGCTGCTCATGAG
CAATGGGAGATGCTGCTGATATTAATGCTAGAGATCTCAATTTGTGC
CGAAGCTCAATGGACTTACAGATTTGCGACCTCTGCTCACTGATGATA
TGATGCTGCTCACTGCTGCTCTAGTGTGATGCTGCTGCTGCTGGA
TGAGCATTGCTGCTGCTGCTGCTCTCAAACTCTTTGCTATGCAAT
GGCATATAGGTTCAATGGCATGGAATACCAAAATGTTCTCTATGAGA
ACCAAAACAAATGCCCAACAAATTTTAAACAGGATGATGCTCAATTCAA
GATCACTTCAACAACTCACTGATGAGGCAAGCTGCAAGAGCTTGT

-continued

TAACCAGATGCTCAAGCATTAACACACTTGTGTAACCACTTAGCTCTA
 ATTTCGTGCAATTTCAAGTGTGCTAAAGATATCTCTTCGGAGCTTGA
 AAGTCGAGCGGAGGTACAAATTGACAGGTTAATTACAGCGAGACTTCA
 AAGCTCTCAAACTATGTAAACACAACTAATCAAGGCTGCTGAAATCA
 GGCTCTGCTAATCTTCTGCTACTAAATGTGCTGAGTGTGTCTTGGG
 CAATCAAAAGAGTGTGCTTTTGTGAAGGGCTACCACTTATGTCTCT
 CCGACCAAGCAGCCCGCATGTGTGTCTCTCTACATGTCACTATGTG
 CATCCGAGGAGGAAGCTCAACCAAGCGCCAGCAAAATTTGTCAAGG
 AAGCATACTCTCCCTGTGAAGGTGTGTTGTGTTAATGGCACTCTTG
 GTTATATACCAAGGAAGCTCTTCTTCTCCAAATTAATTAATACAGACA
 ATCATCTTGTCTCAGGAAATGTGATGTCTTATTGGCATCAATCAAC
 ACAGTTEATGATCTCTGCAAGCTGAGCTGACTCATCAAGAAGAGCT
 GGACAGTACTCTCAAAATCACTACTACCAAGTGTGATCTTGGGACA
 TTTCAGGCATTACGCTCTCTGCTCAACATCTCAAAAGAAATGACGC
 CTCAGTGGCTGTGTAATAATTAAGATCACTATTGACCTCAAGA
 ATTGGGAAATATAGCAATATTAATAAGGCTGTGTATATTTTGGCTG
 GCTTCATGCTGAGCTAATGCCATGCTCATGGTACAACTCTGCTTTGT
 TGATGACTAGTGTGCTGATGCTCAAGGCTGATGCTCTTGTGCTTC
 TCTGCTCAAGTTGATGAGGATGACTCTGAGCAATCTCTCAAGGCTGCTCA
 ATTACATTACACATA

[0030] The S protein has the following amino acid sequence and is referred to herein as SEQ ID NO:23.

MFIFLLFLTLTSGSLDRCTTDFCVQAPNYQHTSHRGVYFDEFESD
 TLVLTQCLFLPFIENVGTHTINHTGHPVLPFGDGIYPAATEKSNVVRG
 NVFGSTNHNKSGVLIINNSTVIRACNPELNDNPFVAVKPMCTQZHT
 MIFDNAPNCTFTYISDAFSLDVSEKGNFKHREFVFNKNDGLFYVYEGY
 QIFDVDDVLPFGWTLKPIFKPLGINITNFRALLTAPSPAQDWGTSA
 AIFVGLKFTPLNLYKDENGITDAVCSQNLAEKLCVKFSFIDRGT
 QTSNFRVPSGDVFRFPNITLCPFGVEPNATKPPSVYAWERKISHCVA
 DYSVLNYSFTFTKCVGSAKTNLDFSPNVDGPPVKGDVDRGIAPG
 QRGVIADYNYKLPDDFGCVLWNTNRINDATSTGNVNYKYRLRHGKLR
 FETDISVFTSPGCKPCTPALNLYCLNDYGYTTTGIGYQPYRVVLS
 FELLNAPATVCGKPLSLDLIXNCVFNPNGLTGTGLVTPSSKRPQFPQ
 FGRDVSDFDTSDVDKPTSEILDISPCPGVSVITPOTNASSAVLVQD
 VHCITDVSTAHADQLTANRYTSTGNVNVQQAQGLIGAEHVDTSEYEDI
 FTGAGICASHYTVLLASTSQSKIVATYMSLGADESSIAYSNWTIAIPWF
 SISITTEVMPVMSHATSDVCHNYICDSETCANLLQYGSFCTQLNRLAS

-continued

GIAAEQRDRTRVEAQVQMYKTPFLYFGGFNSQILPDLKPTKRSFI
 EDLLFHKVTLADAGPMKYGECGLDINARLDICAKQFNLVLPFLTDD
 MIAAYTAALVSGATAGWTGAGAAQIIPFMQNAVRFNGIGVQTQVLYE
 NQKQIANQFNKAIISQIGSLTITTTSTALGKLDVNVNAQALNTLVKQLSS
 NPGAISVLDLILSRDLKVEAEVQIDRLITGLRLSLQTTVQQLRAAEI
 RASANLAATKMSCEVLGSKRVDFCGKGYHLNPFQAFHGVVFLHVTV
 PQERHFTTAPACHEGKAYFPREGVVFNGISWFTIQNFNFSQIITTD
 NTFVSGNCDVVGIIINVTYDPLQPELDSFKELOKYPKNHSTFVDDGLD
 IEGINASVNIQKEIDRLREVAKNLMSLIDQLGLEYEQYIKWPFVYWL
 GF IAGL IAIYVMTILLCCTSCSCLGACSCGSCCKFDEDDSEFVLKGV
 KLMYF

[0031] The S protein can be divided into three structural domains: a large external domain at the N-terminus, a transmembrane domain and a short carboxyterminal cytoplasmic domain. These domains within the S protein of SARS-CoV Urbani strain have been identified using the program TMHMM2.0. (Sonhammer et al. *Proc. Of 6th Int. Conf. On Intelligent Systems for Molecular Biology*. AAAI Press:175-182 (1998). Based on this algorithm, amino acids about 1 to about 1195 comprise an extracellular domain; amino acids about 1196 to about 1218 are part of a transmembrane domain; and amino acids about 1219 to about 1240 comprise the cytoplasmic domain. Removal of residues comprising the transmembrane domain and optionally, the cytoplasmic domain, results in a soluble protein that can be used in the compositions of the invention.

[0032] The large external domain of the S protein is further divided into two sub-domains, S1 and S2. The S1 sub-domain (amino acids about 1 to about 683) includes the N-terminal half of the molecule and forms the globular portion of the spikes. This region contains sequences that are responsible for binding to specific receptors on the membranes of susceptible cells. S1 sequences are variable, containing various degrees of deletion and substitutions in different coronavirus strains or isolates. Mutations in S1 sequences have been associated with altered antigenicity and pathogenicity of the virus. The receptor-binding domain of the S protein of murine hepatitis virus (MHV) is localized within the N-terminal 330 amino acids of the S1 domain. Consequently, the amino acid sequences of the S1 domain may determine the target cell specificity of coronaviruses in animals.

[0033] The S2 sub-domain comprises amino acids about 684 to about 1210 of the S protein. In coronaviruses, the S2 sub-domain of the S protein is usually acylated and contains two heptad repeat motifs. The motifs suggest that this portion of the S protein may assume a coiled-coil structure. The mature S protein forms an oligomer, which is most likely a trimer based on the spike proteins of other coronaviruses. Thus, the S2 subdomain probably constitutes the stalk of the viral spike.

[0034] Non limiting examples of nucleotide sequences encoding the S protein are as follows. It should be noted that S sequences vary between SARS-CoV strains. Virtually any

nucleotide sequence encoding a SARS-CoV S protein is suitable for the present invention. In fact, S polynucleotide sequences included in vaccines and therapeutic formulations of the current invention may change from year to year, depending on the prevalent strain or strains of SARS-CoV.

[0035] From about nucleotide 21492 to about 25080 of the Urbani strain of the SARS-CoV genome encode a soluble extracellular portion of the S protein (Bellini et al. SARS Coronavirus Urbani, complete genome, Genbank accession number AY278741) and has the following sequence, referred to herein as SEQ ID NO: 1:

ATGTTTAACTTTCTATTATTTCTACTCTCACTAGTGGTAAGACCTTGA
CCGGTGCACCACTTTTGATGATGTTCAAGCTCTAAATTACACTCAACATA
CTCATCATAGGGGGGTTACTATCTCTGATGAATTTTGTAGTCAGAC
ACTCTTTTAACTCAGGATTATTTCTTCATTTTATCTCAATGTGTAC
AGGGTTTCATACATAAATCATATCTTTGGCAACCTCTCATACCTTTTA
AGATGGTATTATTTCTGCTGCCACAGGAATCAATGTTGTCGGTGT
TGSGTTTTGGTCTACCATGAACAAAGTCACAGTCGGTGAATTATAT
TAACAACTCTACTAATGTTGTATATACAGCATGTAACCTTGAATGTGTG
ACAAACCTTTCTTTCTGTGTTCTAAACCCATGGGTACACAGCACATACT
ATGATATGATGAATGCAATTAATGTCATTTTCCAGTACATATCTGATGC
CTTTTCGCTGATGTTTCAAGAAATCAGGTAAATTTAAACACTTACGAG
AAGTTGTGTTAAAAAATGAAGAGGGTTTCTCTATGTTTATAAGGGTAT
CAACCTATAGATGATGTTCTGATCTACCTCTCTGTTTTAACACTTTGAA
ACCTATTTTAAAGTGGCTCTGTGATTAACATTACAAATTTAGAGCCA
TTCTTACAGCCTTTTCACTGCTCAAGACATTTGGGGCAGCTCAGCTGCA
GCGTATTTTGTGGCTATTTAAAGCCAACTACATTTATGCTCAAGTATGA
TGAAATAGTACAACTACAGATGCTGTTGATGTTCTCAAAATCCACTTG
CTGAACCTAAATGCTCTGTAAGAGCTTGAGATTGACAAAGGAATTAC
CAGACCTCTAATTTCAAGGTTTGTCTCTCAAGAGATGTTGTGAGATGCC
TAATATTACAACTGTGTCTTTTGGAGAGGTTTAAATGCTACTAAAT
CTCCCTCTGTCTATGATGGAGAGAAAAAATTTCTAATGTGTGCT
GATTACTCTGTCTACAACTCAACATTTTTCACCTTTAAGTGCTA
TGGCTTTCTGCCACTGAATGGTAATGATCTTTGCTCTCCATGTCTATG
CAGATTCTTTTATGTAAGGAGATGATGTAAAGCAATAGCCGCCAGGA
CAAJCTGGTGTATTGCTGATTAATTTAAATTTGGCAGATGATTCAT
GGGTGTGTGCTGCTGGAAGTACTAGCAATGATGCTACTTCAAGTG
GTAAATTAATATAAATATAGTATCTTAGACATGGCAGCTTAGGGCC
TTTGAGAGAGACATATCAATGTGCTCTTCTCCCTGATGGCAAACTGTG
CACCCACCTGCTCTAATTTGTTATTTGGCACTAAATGATTATGTTTTT
ACACCACTATGGCATGGCTACCAACCTTACAGAGTTGTATGACTTTCT
TTTGAACCTTTAAATGACCGCCAGCGTTTGTGACCAAAATATCCAC

-continued

TGACCTTATTAAAGAACGATGTGCAATTTTAAATTTAATGACTCACTG
GTACTGTGTGTTTAACTCCCTCTTCAAAGAGATTTCACCACTTTACAA
TTTGCCCTGTGATTTCTCAATTCACAGTATTCCTGTGAGATCTCAAAAC
ATCTGAATATTAGACATTTCCACTTGTCTCTTTTGGGGGTGAAGTGTAA
TTACACCTGGAAACATGCTTCACTGAAAGTGTCTGTCTATATCAAGAT
GTAACTGCACATGATTTCTACAGCAATTCATGACATCACTCAACACC
AGCTTGGGCATATATTTCTACTGGAACAAATGTATTCAGACATCAAGCAG
GCTGTCTTATAGGAGTCAAGCATGTGACACTCTTATGATGGACATTT
CCATTTGGAGCTGGCAATTTGTCTAGTTAACTACAGTCTCTTATAGAG
TAATCTAGCCAAATATCTATGTGCTTATATCTTCTTAAAGTCTG
ATAGTCAATGCTTACTCTAATAACACCATGCTATATCTACTACTACTTT
TCAATTAGCACTTACACAGAAATATGCTCTTTCTATGCTCAAAACCTC
CGTAGATTGTAAATGTATATCTCGAGATTTCTACTGAATGTGCTAAAT
TGCTCTCAATATGTTGATGCTTTTGCACACATCAATTCAGCTACTCA
GGTATGCTGCTGCAACAGATGCAACACAGCTGAAATGTGCTCAAGT
CAAAACAAATGTACAAACCCCACTTTGAATATTTTGTGGTTTTAAAT
TTTCCAAATATTAACCTGACCTCTAAAGGCCAATGAAGGCTTTTAAAT
GAGGACTGCTCTTAAATAGGTGACACTGCTGATGCTGGCTCATGAA
GCAATATGGCGAATGGCTAGGTATATTAATGCTAGAGATCTCATTTGTG
CGCAGAAATTCATGACATTCACATGTTGGCCACTCTGCTCACTGATGAT
ATGATGCTGCTCACTAGCTGCTGCTAGTTAGTGGTACGCCATGCTG
ATGACATTTGTGCTGCGCGCTGCTCTCAAAATCTTTTGTCTATGCAAA
TGCCATATAGGTTCAATGGCATGGAGTTACCCAAATGTTCTCTATGAG
AACCACAAACAAATGCCCAACCAATTTAACAAAGGGATTAATCAAAATG
AGATCACTTACAAACACATCACTGATTTGGCCAAAGCTGCAAGAGCTG
TTAACAGAAATGCTCAAGCATTAACACATCTGTTAAACCACTATGCTCT
AATTTGTGTCATTTCAAGTGTGCTAAATGATGCTGCTGATGCTGCTGAT
TAAATGCGAGGGAGGTACAAATGACAGGTTAATTAACAGCGAGCTTC
AAGGCTTCAACCTATGTACACACACATTAACAGGCTGCTGGAATAT
AAGGCTTCTGCTAATCTTGTGCTCAATAAATGTCTGATGTGTTCTGG
ACATCAAAAGAGTTGACTTTTGTGGAAGGGCTACCACTTATGCTCT
TCCACAAAGCAGCCCGCATGTTGTGCTCTCTTATGATGCTACGATATG
CCATCCAGAGAGAGAACTTACCACAGCCCGCAGCAATTTGTATGAGG
CAAGCATACTTCCCTGTAAGGCTGTTTGTGTTAATGGCACTTCT
GGTTTATTAACAGAGGAATCTTCTTCTCCAAATATTAATCAACAGAC
AATCATTTGTCTCAGGAATTTGATGTGCTGTTATTTGGCATCAATGAA
CACAGTTATGATGCTCTGCACATCTGAGCTGAGCTCATTCAGGAAGAGC
TGGCAAGTACTTCAAAATATACATACAGCAATGTTGATCTGGCCAC
ATTTACAGGCAATACGCTCTGTGCTCAACATCAAAAGAAATGACCG

—continued

CCCTAATGAGGTCCTCAAAAATTTAAATGAATCACTCATGCACTTCACG
AAATGGGAAATATGACCAATATATTAATGGCCTTG

[0036] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:1, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response. The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0037] The amino acid sequence of the soluble S protein encoded by SEQ ID NO:1 has the following sequence shown below and is referred to herein as SEQ ID NO:2:

MFIFILLFLTLTSGSLDRCTTFDDVQAPNPTQHTSSNRGVYFDFEIRSD
TLYLTLQLFLPPYSNVIGFHTINHTPGNVPVPPKDDIYFAATEKSNVVRG
WVFGSTHNKSGQVILINSTNVIRACNIFLCNDFFAVSEKMGQTQHT
MFDNAPNCTFEYISDAFSLDVS EKSGRFKHLREFVFNKDKFLYVYNGY
QIDVVDLPFGFHTLKPFLKPLGININFRAILTAFSPAQDWGTSAA
AYFVGLKPTTFMLKYDENGTITDAVDCSNPLAEKLSKVSFEIDKGIY
QTSNFRVFSQDVFRFPNTHLCPFGFVFNATKFPFSVTAERKKIENCA
DTSVLNYSSTFFSTKCYOVSAIKLNDLCPFSNVYADSPVVKGDVRQIAFG
QYGVADYNYKLPDPPMGCVLAWNHTIDATSTGNVNYKYRLRHGKLRP
FERDISNVFSPDQKCTPEPALNCVPLMDYGFYTTTGIGTQYRVVLS
PELLNAPATVCGPFLSTDLKNCVNFNFWGLTGTGVLTPSSEKRFQFPQ
FQRDVSPTDSVRPKTSEILDIPSCFSGGVSTPQTWAGSEAVLVQD
VNCTDVSFAHADQLTPAMRIYSGNVNVPQVQAGCLGAERHVDSTECIDI
PIAGQICAEYHTSLSTSQKSVATYHSLGADSEIAYSNTIAIPTN
SISITFEVMPVSHAKTSVDCHNICYGDSSTECANLLQYGSFCTQLNRALS
GLAAEQDRNTRVFAQVQMYKTPTLTYGGFNFSQILPDPLKPTKRSFI
EDLLPRKVTADAGPMKQYGECLGDINARDLCAKPNGLTVPLLLTDD
MLAATTAALVSGTATAGVIFGAGAAQIPFAMQMANRFGNGVGTQWLYE
NQKQIANQPNKALISQIQESLITSTALGDLQDVNNAQALNVLKQLSS
NPGAISVINDILRLDQYAEVQIDRLTLGRILQSLQTVVQQQLIRAAEI
RASANALNMESCEVLQSKRVDPCGKTHLMSFPQAPHGCVFLHVTYV
PQGRNPTTAPACIECHBSKAYFPRGQVFNFGNSWFTIQNRNFSQIITDD
NTFVSGKCDVIGIINNTVYDPLQELDFKELDKYVKNHSTPVDLGD
ISGINASVNIQKSIDRLSEVAKNLHESLDLQELQKTYKIYKFWL

[0038] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S polypeptide comprising an

amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:2, wherein said polypeptide raises a detectable immune response. The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0039] A conserved protein domain program on the National Center for Biotechnology Information's web site (www.ncbi.nlm.nih.gov) was used to predict domains within the SARS-CoV S protein. Two domains, S1 and S2, were predicted within the soluble portion of the S protein. The S1 domain spans from amino acids about 1 to about 683 of the S protein. The nucleotide sequence encoding the soluble S1 domain from SARS-CoV Urbani strain has the following sequence and is referred to herein as SEQ ID NO:3:

ATGTATTATTTCTTATTATTTCTTACTCTCACTAGTGGAATGACCTTGA
CCGGTCACCACCTTTTGATGATGTCTCAAGCTCCTAATACACTCAACATA
CTCATCTATGAGGGGGTTTACTATCTGATGAJATTTTATGATCAGAC
CTCTTTATTAACTACAGATTTATTTCTTCCATTTATCTTAATGTATAC
AGGGTTTCACTACTATTAATCATACAGTTTGGCAACCGCTCATACCTTTTA
AGGATGATATTTATTTCTGCGACAGAGAAATCAATGTTGTCCGTGTG
TGGGTTTTTGGTCTACCAAGCAACAAGTACAGATCGGTGATTATAT
TAACAATCTACTAATGTTGTACAGCAGATATACCTTGAATGTGTG
ACAACCCCTTCTTGTCTGTCTTCAACCACTGGGTACACAGACACATACT
ATGATATTCGTAATGCAITTAATTTGACCTTTTCGATACATATCTGATGC
CTTTCCCTCGATGTTTCAAGAAAGTCAGGTAAATTTAAACACTTACAG
AGTTTGTGTTTAAAAAAGAGTGGGTTCTCTATGTTTATAGGCTGAT
CAACCTATAGTGTAGTGTGATCTACCTCTGTGTTTAAACATCTGAA
ACCTATTTTAAAGTGGCTCTGTGATATTAACATAAAATTTAGAGCCA
TCTCTACAGCCCTTTTCACTGCTCAGACATTTGGGCGACCTCAGCTGCA
GCGTATTTTGTGGCTATTAAAGCCACTACATTTATGCTCAAGATAGA
TGAAAATGCTACACACAGATGCTGTTGATGTCTCTCAAAATCCACTG
CTGACTCAAAATGCTGTTAAAGAGCTTGAAGTACAAAGGAATTCAC
CAGACCTCAATTTAGGTTGTGCTCAGAGATGTTGAGATCTCC
TAATATTACAAACTGTGTCTTGTGGAGAGGTTTAAATGCTACTAAAT
TCCTCTCTCTATGCTAGGAGAGAGAAAAATTTCAATGTGTGTGCT
CTTACTCTGTGCTCTCAACACTCAACATTTTTCCTCAATGTAGTGCTA
CAGATCTTTTGTAGTCAAGGAGATGATGTAGCAAAATAGCCAGAGA
CAAACTGGTGTATGCTGATTAATAATAAATGCCAGATGATTTCAT
GGGTGTGTCTGCTGTGGAATCAAGAACATTTGATGCTACTCACTCACT
GTAAATTAATAATAATAATAGGTATCTTAGACATGCCAAGCTAGGCC
TTTGAGAGAGACATCTAATGTGCTTCTCCCTGATGGCAAACTG
CACCCACCTGCTCTAATGTTTATGAGCAATTAATGATTATGTTT

—continued

ACACCCTACTGGCATTGGCTACCAACCTTACAGAGTGTGACTCTTCT
TTTGAACTTTTAAATGACCGCCGACGGTTTTGGACCAAAATATCCAC
TGACCTATTAGAACCAGTGTGTCAATTPTAAITTTAATGGACTCAGTG
GTACTGTGTGTAACTCTTCTTCAAGAGATTCAACCAATTCAACAA
TTTGCCGTGATGTTCTGATTCTACTGATTCGCTTGAGAGCTCTAAAC
ATCTGAAATATTAGACATTTCACTTGTCTCTTTTGGGGGTGAAGTGTA
TTACACCTGGAAACATGCTTCATCTGAAGTGTCTGTCTATCAAGAT
GTAACTGCACTGATGTTTCTACAGCAATTCATGCAGATCAACTCAACACC
AGCTGGCGCATATATCTACTGGAACAATGTATCCAGACTCAAGCAG
GCTGTCTTATAGAGCTGAGACTGTGCACACTTCTTATGAGTGCACATT
CCTATTGGAGCTGGCAATTTGTGAGTAGTACACAGATTTCTTTATTACG
TAGTACTAGCCAAAATCTATTGTGGCTTATCATATGTCTTTAAGTGTCT

[0040] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S1 polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:3, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response. The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0041] The amino acid sequence of the soluble S1 protein encoded by SEQ ID NO:3 has the following sequence shown below and is referred to herein as SEQ ID NO:4:

MFIFLLFLITSGSLDRCTTFDWDQAPNYQHTSSMRGVYTFEIPRSD
LTYLTQDLFLPFYSNVTGHTINHTPNVPIPKDGIYFAATEKSNVVRG
WVFGSTNNKSGSVIIINHTSVIRACNFELCDNFFAVSKPMQTQHT
MTFDNAPNCTFEYISDAFSLDVSQSNKFKRSEFVKNKFDGLVYVQY
QPIDVVRDLSPFWTLKFIKFLPLGINTHPRAILTAFSPAQDIWUTSAA
ATFVGYLKFPTTHLKYDENGTITADVDCSQFPLAEKCSVKSFIEDNGIY
QTSNFRVPSGDVRFPHNLTCFFGVNPATKFFSVYAWERKKISNCA
DSVLYNSTFFSTFKYGVSAIKNDLFCFKSVIADSPVVKGDVDRQIAPG
QGVGIADNYKLFDDFPGVCLAWNTNIDATSTGNYNKYRLRHGKLRP
FERDISKVFSPDKQKCTPFALNCTHPLMDYGFYTTGIGQYFRRVVLG
PELLNAPATVCGFKLSTDLIKNQCVNPNFQITGCVLTPSKRFPQFPQ
VHCTDVSTAHADGLTPAWRIYSTGNVFNQTAGLIGNEHVDYTESCDI
PIGAGTCASHYVLLSSTSGSKSVIAYTMSLGA

[0042] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S1 polypeptide comprising

an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO:4, wherein said polypeptide raises a detectable immune response. The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0043] The S2 domain spans from amino acids about 684 to about 1210 of the S protein. The nucleotide sequence encoding the soluble S2 domain from SARS-CoV Urbani strain has the following sequence and is referred to herein as SEQ ID NO:5:

GATAGTCAATTGCTACTCTAATAACCACTTGTACTACTACTAACT
TTCAAATGACCTACTACAGAGTAAGTGTCTTCTATGCTAAACCT
CCGTAGATTGTAATATGATCATCTGCGGAGATTCTACTGAATGTGCTAAT
TGTCTTCCANTAGGTGACTTTTGCACACAATAAATGTGCACTCTC
AGGTATTGCTGCTCAACAGAGTGCACACACGTGAAGTGTCTGCTCAAG
TCAACAATGTACAAAACCCCACTTTGAATATTTTGGTGGTTTAAAT
TTTTCAAAATATTACCTACCTTAAAGCCACTAAGAGTCTTTTAT
TGAGACTTGTCTTTAATAGGTGACATGCTGCTGCTGCTCATGA
AGCAATAGGCGAATGCTAGTGTATATTAATGCTAGAGATCTCAITTTG
GCGAGAAGTCAATGACTACAGTGTGCACTGCTGCTGCTGCTGATGA
TAGTATGCTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
GATGGACATTTGGTGTGCGCTGCTCTCAAACTCTTTGCTATGCAA
ATGGCATATAGGTCAATGCAATGGAGTATCCAAAATGTCTCTATGA
GAAACAAAACAAATGCCAACCAATTTAAACAGGCGATTGTCAAAATG
AAGATCACTTACACACATCACTGCAATTTGGCAGCTGCAAGAGCTT
GTTAACAGAAATGCTCAAGCATTAAACACACTTGTAAACAACTTAAGCT
TAATTTTGGCAATTTCAAGTGTGCTAAATGATATCTTTCCGAGCTTG
ATAAGTGGAGGCGAGGTACAAATTTGACAGTTAATACAGGCGAGCTT
CAAAAGCTTCAAACTATATACACACAACTAATGAGGCTGCTGAAT
CAGGCTCTGCTCAATTTGCTGCTACTAAATGTCTGAGTGTGCTTGT
GACATCAAAAGAGTTGACTTTTGGAAAGGGCTACACTTATGCTCC
TPTCCACAGCAGCGCCGAGTGGTGTGCTCTTCACTACTGTCACTATGT
GCATCCAGGAGAGGAACTTCAACACAGCGCCAGCAATTTGTATGATAG
GCAAGCATACTTCCCTGTGAAGGTGTTTGTGTTTAAATGGCACTTCT
TGCTTTATACACAGGAGAACTCTTTCTTCAACAAATTAATCTACAGA
CAATACATTTGTCTCAGGAATTTGATGTGCTGTTATGGCATCAATCA
ACACAGTTTATGATCTCTGCAACCTGAGCTGCACTTCAAGAGAGAG
CTGGACAAGTACTTCAAAATATCATACATCAGCAATGTGTTCTGGCGA
CATTTTCAGGCAATACGCTCTGCTGCTCAACATCAAAAGAAATGAC
GCTCAATGAGTGTGCTCAAAATTTAAATGAACTCACTATGACTTCA
GAATGGGAAATATGACCAATTAATTAATGAGTGTGCTGCTGCTGCTG

[0044] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S2 polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:5, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response. It should be noted that in order to achieve a polynucleotide "operably encoding" a SARS-CoV S2 polypeptide, at least a methionine codon (ATG) would need to be included, in frame, upstream of the polynucleotide presented herein as SEQ ID NO:5. An example of such a polynucleotide includes, but is not limited to the following, presented herein as SEQ ID NO:54.

ATGGATAGTTCAATGCTACTCTAATAACACCATTGCTATACCTACTAA
 CTTTCAATTAGCACTACAGAGTAATGCCCTTTCTATGCGTAAAA
 CCTCCGTAGATTGTATATGTACATCGCGAGATTCTACTGAATGCTCT
 AATTGCTCTCCCAATATGTAGCTTTTGCACACAATACTGTCACAT
 CTCAGTATTGCTGCTGGAACAGGATCGCAACACACGTGAAGTGTGCTC
 AAGTCAACAAATGTACAAAGCCCACTTTGAAATTTTGGTGGTTT
 AATTTCACAAATATCTGACCTCTTAAGGCCAATGAGGCTCTT
 TATTGAGGACTTGCTCTTAAATAGTGACACTCGCTGATGCTGGCTCA
 TGAAGCAATATGGCAATGCTAGGTGATTAATGCTAGAGATCTCAAT
 TGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCTCTGCTCACTGA
 TGATATGATTGCTGCTACTACTGCTCTAGTTAGTGCTGCTGCACTG
 CTGATGGAATTTGTGCTGCTGCTCTCTCAAACTCTTTGCTATG
 CAAATGGCATATAGGTTCAATGGCATGGAATGCTACCCAAATGTTCTCTA
 TGAGAACCAAAACAATGCCAACCAATTTACAAAGGAGATTAGTCAAA
 TCAAGAACTACTTACAAACATCAACTGCTGGCGAAGCTCGAAGAC
 GTTGTTAACCGAATGCTCAAGCATTAACCACTTGTATTAACCACTTAG
 CTCTAATTTTGGTGCATTTCAAGTGTCTAAATGATATCTTTCGCGAC
 TTGATAAGTGGAGCGGAGTACAAATGACAGGTTAATACAGCGAGA
 CTTCAAGCCCTTCAACCTATGTACACACCAACTAATCAGGGCTGCTGA
 AATCAGGGCTCTCTTAATCTGCTGCTACTAAATGTTCTGAGTGTCTC
 TTGACAATCAAAAGAGTTGACTTTTGTGGAAGGCTACCACTTATG
 TCTCTCCACAGCAGCCCGCACTGGTGTGTTCTCTCACTATGCTACGTA
 TGTGCCATCCGAGAGGAAGTCAACACAGCCGCAAGATTGTCATG
 AAGCGAAGCATACTTCCCTCGTGAAGGTGTTTGTGTTTATGCGCACT
 TCTTGTTTATACAGAGGAACTCTCTTTCTCCACAAATAATCTACTG
 AGACAATACATTTGCTCAGGAATATGTATGTCTTATGTCATCATTA
 ACAACAGGTTTATGATGCTCTGCACTGAGCTGCACTCATTCACAAAGA
 GAGCTGCAAGTACTTCAAAATATACATCACCGATGTTGATCTTGG
 CGACATTCAGGCATTAACGCTCTCTGCTCAACTCTCAAAAGAAATG

-continued

ACCGCCTCAATGAGCTGCTAAAAATTAAATGAATCACTTATGACCTT
 CAGAAATGGGAAATATGAGCAATATTAATTAATGAGCGCTTGG

[0045] The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0046] The amino acid sequence of the soluble S2 protein encoded by SEQ ID NO:5 has the following sequence shown below and is referred to herein as SEQ ID NO:6

DSSIAYSNNTIAIPTNFSISITTEVMPVSHAKTGVDCNMHYICGDSIECAN
 LLLQYGSFCTQLNRALSGIAEGQDRNTEVFAVQKMYKPTLKYPGGFN
 FSQILPDPLKPKRSPFIEDLLPNKVTADAGFVKQVGECLGDINARDLIC
 AQKPNGLTVLPELLTDNIAATYALVSGTATAGTGTGAGAAQIPFAHQ
 MAYRPNIGVTVNLYENYQKIQANFNKALSIQISLTTSTALGKLDV
 VHNQAALNTLVKQLSGNFGAISVLDILSLRLOKVEAGVQIDRLITGR
 QSLQTVYVQQLIRAAEIRASANLAATKMSCEVLGSKRVDFCGKGYHMS
 FPQAAPHGVPVFLHVTYVPSQSRNFTTAPACHGKATPFRGQVFNQGT
 WFTIQRNFTSPQIITDNTVSGNCDVIGIINNTYDPLQELDSFKKE
 LDYKFNHNTSPDVLGDSIGINAVVNIQKEDIRLNEVAKNLSLIDQL
 ELGKYEYQIKNFW

[0047] The amino acid sequence of the soluble S2 protein encoded by SEQ ID NO:54 has the following sequence shown below and is referred to herein as SEQ ID NO:56

DSSIAYSNNTIAIPTNFSISITTEVMPVSHAKTGVDCNMHYICGDSIECA
 NLLQYGSFCTQLNRALSGIAEGQDRNTEVFAVQKMYKPTLKYPGGFN
 NFSQILPDPLKPKRSPFIEDLLPNKVTADAGFVKQVGECLGDINARDLIC
 CAQKPNGLTVLPELLTDNIAATYALVSGTATAGTGTGAGAAQIPFAHQ
 QNAYRPNIGVTVNLYENYQKIQANFNKALSIQISLTTSTALGKLDV
 VHNQAALNTLVKQLSGNFGAISVLDILSLRLOKVEAGVQIDRLITGR
 LQSLQTVYVQQLIRAAEIRASANLAATKMSCEVLGSKRVDFCGKGYHMS
 SFPQAAPHGVPVFLHVTYVPSQSRNFTTAPACHGKATPFRGQVFNQGT
 SWFTIQRNFTSPQIITDNTVSGNCDVIGIINNTYDPLQELDSFKKE
 ELDPKFNHNTSPDVLGDSIGINAVVNIQKEDIRLNEVAKNLSLIDQL
 ELGKYEYQIKNFW

[0048] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S2 polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:6, wherein said polypeptide raises a detectable immune response. The present invention is also directed to raising a detectable immune response with or without a wildtype or other secretory leader sequence as described below.

[0049] In one embodiment, soluble S, soluble S1 and soluble S2, described herein, are encoded by a polynucleotide which contains the wild-type S secretory leader peptide sequence. The secretory leader peptide of the S protein in SARS-CoV Urbani strain comprises about the first 13 residues of the protein. Marra et al. The present invention is also directed to raising a detectable immune response with or without amino acids about 1 to about 10, about 1 to about 11, about 1 to about 12, about 1 to about 13, about 1 to about 14, about 1 to about 15, about 1 to about 16, about 1 to about 17, about 1 to about 18, about 1 to about 19, about 1 to about 20, about 1 to about 21, about 1 to about 22, about 1 to about 23, about 1 to about 24, and about 1 to about 25 of the secretory leader peptide sequence.

[0050] In an alternative embodiment, the secretory leader peptide of soluble S, soluble S1 and soluble S2 can be replaced by the secretory leader peptide of human Tissue Plasminogen Activator (TPA). The polynucleotide sequences encoding the various S polypeptides with the TPA secretory leader peptide are shown below. Soluble TPA-S (SEQ ID NO:7)

Soluble TPA-S (SEQ ID NO:7)

ATGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGAGG
AGCTTCCTTTCCGCCAGCGCTAGAGGATCGGGAAGTGACCTTGACCGGT
GCACCACCTTTGATGATGCTCAAGCTCTAATTTACATCAACATATCTCA
TCTATGAGGGGGTTTACTATCTGATGAATTTTATGATCAGACACTCT
TTATTTAACTCAGGATTATTTCTTCCATTTTATCTCAATGTTACAGGGT
TCTATCACTATTATCATACATCTTGCCAGCCCTGTCATACCTTTTAAGGAT
GGTATTATTTGCTGCCACAGAAATCAATGTTTGTGCTGTGTGGGT
TTTTGGTCTACACATGACACACAGTCAAGTCTGGTATTATTATTAACA
ATCTACTAATGTTTATATACAGGACGTAACTTTGAATGTTGACAAAC
CTCTTTCTGCTGTTCTTCAACCCATGGGTGACACAGACATATCATGAT
ATTCGATAATGCATTAAATGCACTTTTCCGATACATATCTGATGCTTTT
CGCTTGATGTTTCAGAAAATGTCAGGTAAATTTTAACTTACAGAGGTTT
GTGTTTAAAATAAAGATCGGTTCTCTATGTTTATAGGGCTATCAACC
TATAGATGATGCTGTGATCTACCTCTGTTTATACACTTTGAACCTCA
TTTTTAAGTTCCTCTGATTTAACTTTAGCAAAATTTAGAGCACTCTT
ACAGCCTTTTCACTGCTCAAGACATTTGGGCGCTGCTGCTGAGCCTA
TTTTGTGCTATTATTAAGCCACATCTTATGCTCAAGTATGATGAAA
ATGTGCAATCAGATGCTGTTGATGTTCTCAAAATCAGCTGCTGAA
CTCAATGCTCTGTTAGAGCTTTGAGTTGTTAGCAAGGAATTTACAGAC
CTCTAATTTACAGGTTGCTTCCCTCAGGAGATGTTGTGAGATTCCTAATA
TTACAAACTTTGTGCTTTTGGAGAGGTTTTTAATGCTCAATTAATCTCT
TCTGTCTATGATGCGAGAGAAAAAATTTCTAATGTTGTGCTGATTA
CTCTGTGCTCACTCAACTCAACTTTTTTCAACCTTTAAGTGTCTGAGCG
TTCTGCGCACTAAGTGTGAATGTTTGGCTCTCAATGCTCATGACAGAT

-continued

CTTTTGTAGTCAAGGAGATGATGTAGACAAATAGSCCGAGACAAC
TGTGTATTGTCGATTATAATATAATATGACAGATGATTCTATGGGTT
GTGTCTGCTTGAATCTAGAACATTTGCTGCTACTCTCACTGCTAAT
TATAATTATAATATAGTATCTTGAACATGCGAAGCTTAGGCCCTTTGA
GAGAGACATATCTAATGTGCTTTCTCCCTGTAGTGCACAACTGACACC
CAGCTGCTCTTAATGTTTATTTGGCCATTAATGATTATGTTTTCACCC
ACTACTGGCATTTGGCTACCAACCTTACAGAGTTGTAGTACTTTCTTTGA
ACTTTAAATGCAAGCGCCAGCGTTTGTGACCAAAATATTCACATGACC
TTATTAAGACACAGTGTGCTAATTTAAATTTAATGAGCTCACTGTGACT
GGTGTGTAACTCTCTTCTCAAGAGATTCTCAACCTATTCAACAAATTTGG
CCGTGATGTTTCTGATTCTACGATTCCTGCTGAGATCTTAACATCTG
AAATATTAGACATTTCACTTCTGCTTTTGGGGGTAAAGTGTAAATACA
CCTGGAACAAATGCTCTATCTGAAGTGTGCTTATATCAAGATGTTAA
CTGACATGATGTTTCTACAGCAATCTCATGAGATCACTCAACACAGCTT
GGCGCATATATCTACTGGAACAAATGTTTCCAGACTCAAGCGCTGT
CTTATAGAGCTGAGCATGTGACACTCTTATGAGTGGAGACTCTCTAT
CTAGCAAAAATCTATTGCTGCTTATCACTGCTCTTATGAGTGTATGAT
TCAATGCTTACTCTAATACCACTGCTATACCTACTAAGTTTTCAT
TAGCATTACTACAGAAATATGCTGTTTCTATGCTTAAACCTGCTGAG
ATTGTAAATATGATCATCTCGGAGAAATCTCATGAATGTGCTAATTTGCT
CTCCAAATATGAGTCTTTTGACACAACTAAATGTGACACTCTCAAGTAT
TGCTGCTGACAGGATCGCAACACATGAAATGTTGCTCAAGTCAAC
AAATGTACAAAACCCCACTTTGAAATTTTGTGTTGTTTAAATTTTCA
CAATATTATCTGACCTCTTAAAGCCACTAGAGGCTTTTATTTAGAGA
CTGTGCTTTAATAGGTGACACTGCTGATGCTGCTCTATGAGCAAT
ATGGGAAATGCTTAGGTATTTAATGCTGAGGATTTAGTAAATCAAGAT
AAGTCTAATGGACTTACAGTGTGACACTCTGCTCACTGATGATGAT
TGCTGCTCACTGCTGCTCTGATGTTGATGCTGCTGCTGCTGATGGA
CATTTGTGCTGGGCTGCTCTCAATACCTTTTGTATGCAATGGCA
TATAGGTTCATAGGCACTGGAGTATCCAAAATGTTCTTATGAGAACCA
AAACAAATGCGCAACCAATTTAACAGCGGATTTAGTAAATCAAGAT
CACTTCAACAACTCACTGCTGTTGGCAGCTGCAAGCGTTGTATAC
CAGATGCTCAAGCAATTAACACACTGTTAACAACATTAGCTCTAATTT
TGGTGCAATTTCAAGTGTGCTAATGATATCTTTTCCGACCTGTGATAAG
TCAAGGCGAGGTCAAAATGACAGGTTAATCAAGCAGCACTCAAAAC
CTTCAAACTATGTAAACAACTAATCAAGGCTGCTGAAATCAGGGC
TTCTGCTAATCTTCTGCTCAATAAATGCTGATGTGTCTTGGCAAT
CAAAAGAGTTGACTTTTGTGGAAGGGCTACCACTTATGCTCTCCCA

-continued

CAAGCAGCCCCGATGGTGTGTCTCTCTCATGATGTCAGGTATGTGCGATC
 CAGGAGGAGAACTTCACACAGGCCACAAATTTGTCATGAGGCCAAG
 CATACTTCCTGTGAAGTGTGTTTGTGTTAATGACACATCTCTGGTTT
 ATTACACAGAGAACTCTTTCTTCCACAAATAATACACAGACAATAC
 ATTTGTCTCAGGAAATGTGATGCTGTATATGGCATCATTAACAAACAG
 TTTATGATCTCTGCAACCTGAGCTGACATTCACAAAGAGAGCTGAGC
 AAGTACTTCAAAAATCATACATACACAGATGTTGATCTTGGGAGCATTC
 AGGCATTAAAGCTCTCTGTCTGCAACATTCACAAAGAAATGACGGCTCA
 ATGAGTCTGCTAAAAATTTAAATGAATCATCTATTGACCTTCAAGAATTG
 GGAAAAATGAGCAATATTAAATGGCGCTGG

Soluble TPA-51

(SEQ ID NO:9)

ATGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGC
 AGCTTCGTGTTGCCGAGCGTAGAGATCGGGAGCTGACCTTGACCGGT
 GCACACATTTTGATGATGTCTAAGCTCTCAATACATCAACATCTTCA
 CTATAGGGGGGTGTTATCTATCTGATGAATTTTATAGATCAGACACTCT
 TTATTTAACTCAGGATTTATCTTCCATTTATCTTAATGTTTACAGGT
 TGCATCTATTATATCATCTGTTGGCAACCTGTCTACACTTTTAAGAT
 GGTATTTATTTGCTGCCACAGAGAAATCAAAATGTTGCTGCTGGTGGGT
 TTTTGGTCTACACATGAACAAAGCTCAGCTGGGTATTTATTAAACA
 ATTCATCTAATGTTGTATACAGGACATGTAACTTTGAATGTGTGACAC
 CTTTCTTTGCTGTTCTTAAACCCATGGGTACACAGACATATCTATGAT
 ATTCGATTAATGATTTAATGCTACTTCTGAGTACATCTGATGCTTTT
 CGCTGTGATGTTCAAGAAATGTCAGGTAAATTTAAACACTTACAGAGATT
 GTGTTTAAAAATAAGATGGGTTTCTCTATGTTTAAAGGGCTATCAACC
 ATATGATGTAGTCTGTGATCTACCTCTGTTGTTTAAACATTTGAACCTA
 TTTTAAATGCTCTCTGTTGATTAACATTAACAAATTTAGAGCCATCTT
 ACAGCCTTTTACCTGCTCAAGGATTTGGGACACCTAGCTGACGCTA
 TTTTGTGCTATTTAAAGCCACTCATTTATGCTCAAGTATGATGA
 ATGGTCAATCAAGATGCTGTTGATGTTCTTCAAAATCCACTGCTGCA
 CTCAAATGCTCTGTAAAGAGCTTGAAGTATGACAAAGGAATTTACAGAC
 CTCATTTACAGGTTGTCTCTCAGGAGATGTTGTGATGTTCTCTAATA
 TCTCAACCTTGTGCTCTTTGAGAGGTTTAAATGCTACTAAATCCCT
 TGTGCTCATGATGAGGAGAAAAAAATTTTAAATGCTGCTGATTA
 CTGTGTGCTCACTACACTCAACATTTTCTCAACCTTTAAGTGTATGCG
 TTTCTGCCACTAAGTGAATGATCTTCTCTCAATGTCTATGAGAT
 TCTTTGTAGTCAAGGAGATGATGTAAACAAATAGCGCCAGGACAAAC
 TGTGTTATGCTGATTAATTAATAATGCGAGATGATTTGATGGGT
 GTGCTCTGCTGCAATACAGGAGATGATGCTACTTCACTGATTAAT

-continued

TATAATTATAAATATAGGTATCTTGACATGCGCAAGCTTAGGCGCTTTGA
 GAGAGACATATCTAATGTGCTTCTTCCCTGATGGCAACCTTGCACCC
 CACTGCTCTTAATGTGTATTTGGCCATTAAGATTTGTTTAAACAC
 ACTACTGGCATGCTACCAACCTTACAGAGTGTAGTACTTTCTTTGA
 ACTTTTAAATGCAACGGCCAGGTTTGTGGACAAATATATCCATGAC
 TTATTAAGAACAGGTGTGCAATTTAATTTAATGGACTCACTGTACT
 GGTGTGTTAACTGCTCTTCAAGAGATTTCAACATTTCAACAATTTGG
 CCGTGTGTTTCTGATTTCACTAGTATCCCTGTGAGATCTTAAACATCTG
 AAATATTAGACATTTCACTTGTCTTTTGGGGTGTAGGTGATTAACA
 CCTGGACAAATCTTCACTGAAATGTCTGCTTATATCAAGATGTAA
 CTGCACTGATGTTTCTACAGCAATCTACAGATCACTACACACAGCTT
 GCGCATATATTTACTGCGAAACATGATTTCAAGACTCAAGCAGGCTGT
 CTATAGAGAGCTGAGCATGTGACACTTCTTAAGTGGAGCATCTTCTAT
 TGAGCTGGCATTTGTGCTAGTTTACCATACATTTCTTTTAAATGATGA
 CTAGCCAAAAATCTATGCTGCTTATCATGTCTTTTAGTGGC

Soluble TPA-52

(SEQ ID NO:11)

ATGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGGAGC
 AGCTTCGTGTTGCCGAGCGTAGAGATCGGGAGATGTCATCTGCTT
 ACTCTAATTAACACATGCTATACCTTAACTTTTCAATTAAGCATTAAT
 ACAGATTAATGCTGCTTCTATGCTTAAACCTGCTGATGATGATAT
 GTACATCTGCGGAGATTTCTATGATGCTGATTAATTTGCTTCTCAATAG
 GTAGCTTTTGCACACAACTAAATGTGCACTCTCAAGTATGCTGCTGA
 CAAGATCGCAACACAGCTGAGTGTGCTGCTCAAGTCAACAAATGATCA
 AACCCCACTTGAATATTTTGTGTTTAAATTTTCAACAATATTAC
 CTGACCTCTAAAGCCCACTAAGAGTCTTTTATTAAGGACTTGTCTT
 AATAGGTGACACTGCTGCTGATGCTGCTTCAATGAGCAATGCGGAAT
 CTTAGTGATTAATGCTGAGATCACTATTGCTGCGGAGATGCTCAATG
 GACTACAGTGTGCTGCTGCTGCTGCTGATGATGATGCTGCTGCTAC
 ACTGCTGCTTAGTGTAGTGTGCTGCTGCTGCTGATGATGATGCTGCTG
 TGCGCTGCTCTTCAAAATACCTTTGCTGCTGCAAAATGCTATAGTCTCA
 ATGCGCTGAGGTATCCCAAAATGCTCTTATGAGAACCAAAACAAATC
 GCAACCAATTTAACAAGGAGATGATCAAAATTCGAAGATCACTAACAC
 AACATCAACTGCTATGCGGAGCTGCAAGAGTGTGTAACCAAGATGCTC
 AAGCATTAACACACTTGTGTAACCACTTAGCTTAATTTTGTGCAAT
 TCAAGTGTGCTAAATGATATCTTTGCGACTTGATAAAGTGGAGGCGA
 GGTAACAATGACAGTAAATACAGGAGACTTCAAGGCTTCAAGCT
 ATGTAACACAACTAATCAGGCTGCTGAATCAGGCTCTGCTAAT
 CTGCTGCTACTAAATGTCTGATGCTGCTTGTGACAACTCAAAAGAT
 TGACTTTGTGGAAGGCTACCACTTATGCTCTTCCACAAAGAGGCC

-continued

CGCATGGTGGTTGCTCTCTCATATGTCACGTATGTCACATCCAGGAGAG
 AACTTCACACAGCGCCGCAATTTGTCATGAGGCAAGCACTACTTCCC
 TCGTGAAGGTGTTTGTGTTAATGGACCTCTCTGTTTATTACACAGA
 GGAACCTCTTTCTTCCACAAATAATTACTACAGACATATCATTTGTCTCA
 GGAATATGTGATGCTGTTATTTGGCATCTATTAAACACAGCTTATGATCC
 TCGTCAACCTGAGCTGACTCTATTCAAGAGAGCTGGACAGTACTTCA
 AAAATCATACATCAACAGATGTTGATCTCTGGCAGCATTTACGACATTAA
 GCTTCTGTGTCACAACTCAAAAGAAATTCAGCCCTCAATGAGGTGCG
 TAAAGAAATTAATGATCACTACTGACCTTACAGAAATGGGAAATATG
 AGCAATATATTAATGGCTTGG

[0051] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S, S1, or S2 polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NOs:7, 9, or 11, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response.

[0052] The amino acid sequences of the soluble S protein, S1 and S2 proteins with the TPA secretory leader peptide are shown below. Soluble TPA-S protein (SEQ ID NO:8)

Soluble TPA-S

(SEQ ID NO:8)

MDAMKRGCLCCVLLCGAVFVSPARGSGSDLRCTTFDQVAPNYTQTS
 SHRGVIYPDEIFRSGLYLTLQDLFLFFYSNVTGFHTINHTGRVPIPFKD
 GIFYAATEKSNVVRGWFSGTSHNKSQSVIIINNSHNVRACNFELCDN
 PFFAVSKPMGTQTHMTIFDPAFCTFEYISDAFSLDVEKSGNFMELREF
 VFKNDGFLIVYKGTQIDVVRDLSPGFTLKPFLKPLGININTFRAIL
 TAFSPAQDWTGSAAYFVGLKPTFFPKLYDENGITDADVCSQNFAL
 LKCSVKSEIDKGIYQTSNFRVPSGDVVRPNITLCPFGVEFNATKFP
 SVYANERKKISNCVADSVLYNSTFTSTFKCYGVSATKLNLCFSNVYAD
 SPVVKGDVGRQIAPQGTGIADYNYKLPDFFGCVGLMWNTRINDATSTGN
 YNYKYRLRHGKLAPFERIDSNVFFSPDGKCTFPALNCYWFLNDYGFIT
 TTGIGYQYRVRVLSPELLNAPATVCGPKLSTDLIKMQCVNFMNGLTGT
 GVLTSPSKRPFQPGQGRDVSDFTSVDRPNTSEILDSPSCPGGVSVIT
 PGTNASSEVAVLYQDVNCTDVSATIAHDLQTPANRISTGNVNTQTQAGC
 LIGAHEVDTSEYCDIPGAGICASYHTVLLRSTSGNSIVATYMSLGA
 Soluble TPA-S2 protein
 (SEQ ID NO:12)
 MDAMKRGCLCCVLLCGAVFVSPARGSGDSIAYSHNTIAPIPTFTSIT
 TEVHPFSNAKTSVDCNMYICGDSCECANLLQYGSCTQLNRALSGLIAE
 QDRNREVAQVQKMYKTEPLTKYFGFGNFSQILPOPLKPTKRSFIEDLL
 NKVTLADAGFHKYGECLGDINARDLICAFKFNGLVLPFLLEDMDIAAY
 TAALVSGTATAGTWFGAGAAQIFPANAQYRFGNIGVGTQVLYEKQKQI
 ANQPKKAIQSQISLTTSTALGLQDVVFNQANALYVVKLSSNFGAI
 SEVLNDILSRKLKVEAEVQIDRLITGRQLSITQYTVQLIRAAEIRASAN
 LAATPMSECVLQSKREVDPGCGKHLSMPQAAPAGVTVFLHVTVPSEQR
 NPTTAPALCHGKAYFREGVFPVFNSTNFIQNRNFSFQIITDNTFVS
 GRCDDVIGIINNTVDFLPQLDSEKLELKYFNHSTSPDVLGDISGIN
 ASVNNIQRIDRLNEKVNKLNESLIDLQELKYEQIKNFW

-continued

YRPMGIGVTVNLYENQKQIANQPKAISQDSLETTSTALGLQDVVN
 QHAQALNLTLYKQSSNFHAISSVLNDILSRKLKVEAEVQIDRLITGRQ
 LQYTVFQQLIRAAEIRASANLAATPMSECVLQSKREVDPGCGKHLSMP
 QAAPHGVVFLHVTVPSEQRNFTAPALCHGKAYFREGVFPVFNSTNFI
 ITQRNFFSQIITDNTFVSGNDVIGIINNTVDFLPQLDSEKLEL
 KYFNHSTSPDVLGDISGINASVNNIQRIDRLNEKVNKLNESLIDLQEL
 GRYEQIKNFW
 Soluble TPA-S1 protein
 (SEQ ID NO:10)
 MDAMKRGCLCCVLLCGAVFVSPARGSGSDLRCTTFDQVAPNYTQTS
 SHRGVIYPDEIFRSGLYLTLQDLFLFFYSNVTGFHTINHTGRVPIPFKD
 GIFYAATEKSNVVRGWFSGTSHNKSQSVIIINNSHNVRACNFELCDN
 PFFAVSKPMGTQTHMTIFDPAFCTFEYISDAFSLDVEKSGNFMELREF
 VFKNDGFLIVYKGTQIDVVRDLSPGFTLKPFLKPLGININTFRAIL
 TAFSPAQDWTGSAAYFVGLKPTFFPKLYDENGITDADVCSQNFAL
 LKCSVKSEIDKGIYQTSNFRVPSGDVVRPNITLCPFGVEFNATKFP
 SVYANERKKISNCVADSVLYNSTFTSTFKCYGVSATKLNLCFSNVYAD
 SPVVKGDVGRQIAPQGTGIADYNYKLPDFFGCVGLMWNTRINDATSTGN
 YNYKYRLRHGKLAPFERIDSNVFFSPDGKCTFPALNCYWFLNDYGFIT
 TTGIGYQYRVRVLSPELLNAPATVCGPKLSTDLIKMQCVNFMNGLTGT
 GVLTSPSKRPFQPGQGRDVSDFTSVDRPNTSEILDSPSCPGGVSVIT
 PGTNASSEVAVLYQDVNCTDVSATIAHDLQTPANRISTGNVNTQTQAGC
 LIGAHEVDTSEYCDIPGAGICASYHTVLLRSTSGNSIVATYMSLGA
 Soluble TPA-S2 protein
 (SEQ ID NO:12)
 MDAMKRGCLCCVLLCGAVFVSPARGSGDSIAYSHNTIAPIPTFTSIT
 TEVHPFSNAKTSVDCNMYICGDSCECANLLQYGSCTQLNRALSGLIAE
 QDRNREVAQVQKMYKTEPLTKYFGFGNFSQILPOPLKPTKRSFIEDLL
 NKVTLADAGFHKYGECLGDINARDLICAFKFNGLVLPFLLEDMDIAAY
 TAALVSGTATAGTWFGAGAAQIFPANAQYRFGNIGVGTQVLYEKQKQI
 ANQPKKAIQSQISLTTSTALGLQDVVFNQANALYVVKLSSNFGAI
 SEVLNDILSRKLKVEAEVQIDRLITGRQLSITQYTVQLIRAAEIRASAN
 LAATPMSECVLQSKREVDPGCGKHLSMPQAAPAGVTVFLHVTVPSEQR
 NPTTAPALCHGKAYFREGVFPVFNSTNFIQNRNFSFQIITDNTFVS
 GRCDDVIGIINNTVDFLPQLDSEKLELKYFNHSTSPDVLGDISGIN
 ASVNNIQRIDRLNEKVNKLNESLIDLQELKYEQIKNFW

[0053] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV S, S1, or S2 polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NOs:8, 10, or 12, wherein said polypeptide raises a detectable immune response.

[0054] In a further embodiment, the present invention provides for methods for raising a detectable immune response to the SARS-CoV polypeptides, comprising administering to a vertebrate a polynucleotide which operably encodes polypeptides, fragments, variants, or derivatives thereof as described above.

[0055] The S protein of some coronaviruses contain an Fc-like domain that binds immunoglobulin. Data from the FIPV immunization suggests that high levels of potentially neutralizing antibody may be bound by the Fc-mimicking region of the S protein. Scott, F. W. *Adv. Vet. Med.* 41: 347-58 (1999). Thus, modification or deletion of an Fc region of the SARS-CoV S protein may be useful in the compositions of the present invention.

[0056] The nucleocapsid protein (N) is encoded by about nucleotides 28120 through about 29388 of the Urbani strain of SARS-CoV. (Bellini et al. SARS Coronavirus Urbani, complete genome. GenBank Accession No. AY278741).

[0057] The protein is a phosphoprotein of 50 to 60 kd that interacts with viral genomic RNA to form the viral nucleocapsid. N has three relatively conserved structural domains, including an RNA-binding domain in the middle that binds to the leader sequence of viral RNA. N protein in the viral nucleocapsid further interacts with the membrane protein (M), leading to the formation of virus particles. N is also suggested to play a role in viral RNA synthesis, by a study in which an antibody directed against N inhibited an in vitro coronavirus RNA polymerase reaction. Marra et al. N protein also binds to cellular membranes and phospholipids, a property that may help to facilitate both virus assembly and formation of RNA replication complexes.

[0058] From about nucleotides 28120 to about 29388 of the Urbani strain of the SARS-CoV genome encode the N protein. (Bellini et al. SARS Coronavirus Urbani, complete genome. GenBank Accession No. AY278741) and has the following sequence, referred to herein as SEQ ID NO:13:

```

ATGTCTGATAATGGACCCCAATCAACCAACGATAGTCCGCCCGCATTC
ATTGGTGACCCACAGATTCACCTAGCAATACCAAGATGAGGACCA
ATGGGCGAAGCCAAACAGGCGCCAGCCACCAAGGTTTACCAACAATATC
GGGTCTGGTTCACAGCTCTCACTCAGCATGCGAAGAGGAAGTATGAT
CCCTCGAGGCGCAGGGCGTCCAACTAACCAACATAGTGTCACGATGACC
AAATGGCTACTACGAGAGGATCCCGACGAGTCTCGTGGTGACCGC
AAATGAAGAGCTCAGCCCGAGATGTACTTCTATTAAGTACGAGTGG
CCGAGAGCTCCTCTCCCTAGCGGCTCAACCAAGAGCATCTGATGG
TTGCACTGAGGGAGCTGATATACACCAAGACACATTTGGCACCOCG
AATCTAATAACAATCTCGCACCTGCTCACTCTCTCAAGGACAAAC
ATTTGCCAAAGGCTTCTACGAGAGGAGGAGGAGGCGGCGATCANGCT
CTTCTGCTCTCATCAGCATGATCGGGTAACTCAAGAATCACTCATCT
GGCAGCATAGGGGAAATCTCTCTGCTGAAGTGGCTAGCGAGGTGGTA
AACTGCGCTCGCGTATTTGCTGTAGACAGATGAACAGCTCTGAGAGCA
AAATTCTCTGTAAGGCGCAACAAACAGGCGCAACTGCTCAAGAA

```

-continued

```

TCTGCTGCTGAGGCATCTAAAAAGGCTCGCCAAAAGTACTGCCACAA
ACAGTACACGCTACTCAAGCATTTGGGAGACGTGGTCAGAAACAAACC
ANGGAAATTCGGGGACGAGACCTAATCAGCAAGGAAGTATGATACAA
CATTTGGGCGCAATTCACAAATTTGCTCAAGTGCTCTGCTCTTTGG
AATGTCAGCATTTGGATGGAAGTACACATTCGCGAACATGGCTGACT
ATCATGGAGCCATTAATTTGGATGCAAGATCCCAATCAAGACAC
GTCTACTGCTGACCAAGCACATTTGACCATCAAAACATTTCCACACAC
AGAGCTTAAAGGACAAAAGAAAGAGACTGTGAAGCTCAGCGTTTGC
CCGAGAGCAAAAGAGACGCCACTGTGCTCTCTCTCTCGCGCTGAC
ATGATGATTTTCCAGACACTCAAAATTCATGAGTGGAGCTTCTG
TGATTAACCTCAGGACATA

```

[0059] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV N, polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:13, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response.

[0060] The amino acid sequence of the N protein encoded by SEQ ID NO:13 has the following sequence shown below and is referred to herein as SEQ ID NO:14

```

MSIDHGPQSRQAPRITFGPTDSTDNQNGRIGAPKQRRPGLPNT
ASWFTALTQHGEELRFFRGQGVINTNSGDDIGYRYRATRVRGGR
KMKELSPWFFYLLGTGPEASLPYGANKEGIVNVATGALMTPEKHIGTR
NPRNNAATVTLQPLPGTTLPRGFYARGSRGGSQASRRSSSSRGRSNTSP
GSRGNSPARGASGGSTALALLLLRLAQLSKVSRGQGGQGGQVTVTK
SAJASKKPRKRATKQVHVYQAFRRRQPTQHTFGDQLRQDTYK
HWPQAFAPASAFPMGRIGHEVTPSTWTLTHAIGLTKDKDPQFDN
VILLNKHIDAYETFPPTPEKKKKKIDEAQLPQRQKQPTVTLPLAAD
MEDFSRQLNGMSGASADTQA

```

[0061] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV N polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:14, wherein said polypeptide raises a detectable immune response.

[0062] The N protein contains a nuclear localization sequence (NLS) which directs the protein to the nucleus infected cells or cells in which the protein is expressed. The sequence of the NLS is KTFPPTEPKDKKKITDEAQ (underlined above) and is referred to herein as SEQ ID NO:17. For purposes of the invention, the NLS may be deleted from the protein to obtain a non-nuclear localized version of the protein. The nucleotide sequence of an N

protein lacking the NLS is referred to herein as SEQ ID NO:15 and is shown below.

ATGTCTGATATGACCCCAATCAACCAAGTAGTCCGCCCGCATTAC
 ATTTGGTGACCCACCAATTCACCTGACATAACCAAGATGGAGACGCA
 ATGGGGCAAGCCCAACACAGCGCCACCCCAAGGTTTACCAATAATAC
 CGCTCTGGTTCACAGCTTCACTCAGCATGGCAAGAGGAACCTAGATT
 CCGCTGAGGCGCAGGCGGTTCCAAATCAACCAATAGTGTCCAGATGAC
 AATTTGGCTACTACGAGAGCTACCCGACAGTTCGTGGTGGTACGGC
 AAAATGAAGAGCTCAGCCCGAGATGGTACTTCTATTACCTAGGAAGTGG
 CCGAGAGCTTCACTTCCCTACGGCGCTAACAAAGAGGACATCGTATGGG
 TTGCACTGAGGGAGCGTTGAATACACCAAGACCAATTTGGCACCOC
 AATCCTAATAACATGCTGCCACCGTGCTCACTTCTCAGGAACAC
 ATTGCCAAAGGCTTACGAGAGGGAGGAGGCGGCGCTCAAGCTT
 CTCTCGCTCTCATCAGTATGCGGTAATCAAGAAATCTCACTCT
 GGCACAGTAGGGGAAATTCCTGCTGCAATGGCTGGAGGAGTGGTA
 AACTGCGCTCGCGGATTTGCTGCTAGACAGATGAACAGCTTGAGACA
 AAGTTTCTGTAAGGCGCAACACAGAGCGAACTGTCACTAGAAA
 TCTGCTCTGAGGCACTCAAAAGCGTCCCAAAAGCTACTGCCACAAA
 AAGATCAACGCTCACTCAAGCTTTGGGAGCGTGGTCAGAAACAAAC
 CAGGAAATTTGGGGAGCAGCACTCAATCAAGCAAGGAATGATACAAA
 CATTTGGCGCAAAATGTCACAAATTTGCTCAAGTGGCTGTGATCTTTGG
 AATGTACGCAATTTGGCATGGAAGTCAACCTTCGGGACATGGCTGACTT
 ATCATGGGCACTTAATTTGGATGACAAAGATCCAAATTCAGAACAC
 GTCACTACTGCTGAACAGCACTTACGACATCCCTTTCGCGAGAGACA
 AAGAGAGCAGCCACTGTGACTTCTTCTGCTGCGTGACATGATGATT
 TCTCCAGACACTTCAAAATTCATGAGTGGAGCTTCTGCTGATCAACT
 CAGGCATAA

[0063] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV N polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:15, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response.

[0064] The amino acid sequence of the N protein without the NLS sequence is encoded by SEQ ID NO:15 has the following sequence shown below and is referred to herein as SEQ ID NO:16:

MSDNGPQNSQRSAPRTTGGDTSDTNNQNGRNGARPKRRPQGLPNT
 ASHPTALTQHGXELRFRGQGVPIINRSGDDDTGYRRATRRVRGDD
 KMKELSPRYFTYLLTGTPASLPTGANKGIVVATGALNTPKDIOTR

-continued

NPNNAAATVQLPGQTTLPKGYFARGSRGSGSSRRSSRRSGNSHSTP
 GSRGNSPAMASGGGTALALLLLRLQLSEVSGKQQQQQVTVTKX
 SAARASKKPRKRTATQYNTVQAFRRGPEQTQNTGGDQLIRQDTDK
 HWPQIAPFASASAFFGMSRIGHEVTPSTWLTTHAIGLTKDDKDFPFKN
 VILLNKHIDATPLPQRKEQFTVTLPAADMDFSLQNSHMSGASDST
 QA

[0065] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV N polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:16, wherein said polypeptide raises a detectable immune response.

[0066] The membrane glycoprotein (M) is encoded by about nucleotides 26398 to about 27063 of the Urbani strain of SARS-CoV. (Bellini et al. SARS Coronavirus Urbani, complete genome. GenBank Accession No. AY278741). The M protein differs from other coronavirus glycoproteins in that only a short amino terminal domain of M is exposed on the exterior of the viral envelope. This domain is followed by a triple-membrane-spanning domain, an α -helical domain, and a large carboxy/terminal domain inside the viral envelope. In some coronaviruses, such as transmissible gastroenteritis coronavirus (TGEV), the carboxy/terminus of the M protein is exposed on the virion surface. Glycosylation of the aminoterminal domain is O-linked for MHV and N-linked for infectious bronchitis virus (IBV) and TGEV. Monoclonal antibodies against the external domain of M neutralize viral infectivity, but only in the presence of complement. M proteins of some coronaviruses can induce interferon- α . The M proteins are targeted to the Golgi apparatus and not transported to the plasma membrane. In TGEV and MHV virions, the M glycoprotein is present not only in the viral envelope but also in the internal core structure. (*Field's Virology*, B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman, and S. E. Straus, eds., 4th Edition. Lippincott-Raven, Philadelphia, Pa.).

[0067] From about nucleotides 26398 to about 27063 of the Urbani strain of the SARS-CoV genome encode the M protein, Bellini et al. SARS Coronavirus Urbani, complete genome, GenBank Accession No. AY27874, and has the following sequence, referred to herein as SEQ ID NO:18:

ATGCGACAGCAAGGTACTATTACCGTTGAGGAGCTTAAACACTCTGGC
 ACAATGGACCTAGTAATAGTTTCTTATCTAGCTGGATTATGTGTA
 TACAATTTGCTTCTTAATCGGAACAGGTTTGTGACATATAAAGCTT
 GTTTTCTCTGCGCTTGTGGCCATTAACACTTCTGTTTGTGCTGTC
 TCGCTCTCAGAAATTAATGGGTGACTGGCGGATTCGATTCAATGAT
 CTGTATGTTAGGCTGTGATGTGGCTAGCTACTCTTGTGCTCTTCAGG
 CTGTGTCGTGACCGCTCAATGTGCTCAATCAACCAAGAAACAACT

-continued

TCTTCTCAATGTGCCTCGGGGACAAATGTGGACGAGCCGCTATG
 AAGTGAACCTGTCTATGGTGTCTGTGATCAATCTGTGTCTACTTCGGAATG
 GCCCGACACCCCTGAGGCGGTGACATTAAGACCTCTCCAAAGAGAT
 CACTGTGGTCACTCAGAACGCTTCTTATTACAAATAGGAGCGTGC
 ACGTGTGAGGCACTGATTCAGGTTTGTGTGCATACCAACGCTACCGEAT
 GGAACATAAATAAATACAGACCACGCCGGTAGCAACGCAATATTC
 TTGCTGTAGTACAGTA

[0068] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV M₁ polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:18, or a codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response.

[0069] The amino acid sequence of the M protein encoded by SEQ ID NO: 18 has the following sequence shown below and is referred to herein as SEQ ID NO: 19:

MADNGTIVTELKQLLEQNVLVIGFLFLAWIMLQFATSNRNFLYIKL
 VFLMLNFWTLPCFLVAAYRINWVGIIAMACIVGLMNLSTPVASFR
 LFARTRMNSFWPETHILLNPLRGITIVRPLMESELVIGAVIRGHRM
 AGHPLGRCDIKDLKXITVATSRLLSTYKLGAQRVUTDSGFAAYNRYRI
 GYKLNTRDAGENDNIALVQ

[0070] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV M₂ polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:19 wherein said polypeptide raises a detectable immune response.

[0071] The small envelope protein (E) is encoded by about nucleotide 26117 to about 26347 of the Urbani strain of SARS-CoV (Bellini et al. SARS Coronavirus Urbani, complete genome, GenBank Accession No. AY278741), and has the following sequence, referred to herein as SEQ ID NO: 20:

ATGTACTATTCTGTTTCGAGAGAAACAGGTACGTTAATGTTAATAGCT
 ACCTCTTTTCTTCTCTCTGTGATTTCTTCTGTAGTACACACTAGCCATCC
 TTACTGCGCTCGATTGTGTGCTGACTCTGCAATATTGTAACTGAGT
 TTGATTAACCAACGGTTTACGTCTACTGTGGTGTAAATCTGAATCTC
 TTCTGAAGGAGTTCGTATCTTCTGTCTAA

[0072] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV E polypeptide, wherein said polynucleotide is 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:20, or a

codon-optimized version as described below, and wherein said polynucleotide encodes a polypeptide that elicits a detectable immune response

[0073] Based on protein comparisons with other coronaviruses, the SARS-CoV E protein shares conserved sequences with TGEV and MHV. For some coronaviruses, such as TGEV, the E protein is necessary for replication of the virus, while for others, such as MHV, loss of the E protein merely reduces virus replication without eliminating it completely. Marra et al. The protein sequence is shown below and referred to, herein as SEQ ID NO:21.

MTSPVSEETGLIVNSVLLFLAFVFLVLLTLLAILTALRLCATCCNIVNS
 LVKPTVVIVSRVKMLNSEEQVGLLV

[0074] In a further embodiment the methods of the present invention provide for administering a polynucleotide which operably encodes a SARS-CoV E polypeptide comprising an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO:21 wherein said polypeptide raises a detectable immune response.

[0075] It should be noted that nucleotide sequences encoding various SARS-CoV polypeptides may vary between SARS-CoV strains. Virtually any nucleotide sequence encoding a SARS-CoV protein is suitable for the present invention. In fact, polynucleotide sequences included in vaccines and therapeutic formulations of the current invention may change from year to year, depending on the prevalent strain or strains of SARS-CoV.

[0076] Further examples of SARS-CoV polypeptides within the scope of the invention are multimerized fragments of SARS-CoV polypeptides and polynucleotides that encode multimerized fragments of SARS-CoV polypeptides. The polypeptide fragments of the invention contain at least one antigenic region. The SARS-CoV polypeptide fragments are fused to small assembly polypeptides. Non-limiting examples within the scope of the invention include coiled-coiled structures such as: an amphipathic helix, the yeast CGN4 leucine zipper, the human p53 tetramerization domain, and synthetic coil polypeptides. The SARS-CoV and assembly peptide fusion proteins self-assemble into stable multimers forming dimers, trimers, tetramers, and higher order multimers depending on the interacting amino acid residues. These multimerized SARS-CoV polypeptide fragments have increased local epitope valency which functions to more efficiently activate B lymphocytes, thereby producing a more robust immune response. Also within the scope of the invention are multimerized SARS-CoV polypeptide fragments that maintain conformational neutralizing epitopes.

[0077] Also within the scope of the present invention are combinations of SARS-CoV polypeptides and polynucleotides that encode SARS-CoV polypeptides, where the polypeptides assemble into virus-like particles (VLP). One such combination is, but is not limited to a combination of SARS-CoV S₁, M₁ and E polypeptides or fragments, variants, or derivatives thereof, and polynucleotides encoding SARS-CoV S₁, M₁ and E polypeptides or fragments, variants, or derivatives thereof. Combinations of SARS-CoV polypeptides that form VLPs may be useful in enhancing immuno-

genicity of SARS-CoV polypeptides and in eliciting a detectable immune response to the SARS-CoV virus. Also within the scope of the present invention are methods of producing SARS-CoV VLPs in vitro by using protocols that are well known in the art. The production of VLPs may be performed in any tissue culture cell line that can tolerate expression of SARS-CoV polypeptide. Examples of cell lines include, but are not limited to, fungal cells, including yeast cells such as *Saccharomyces* spp. cells; insect cells such as *Drosophila* S2, *Spodoptera* Sf9 or Sf21 cells and *Trichoplusia* High-Five cells; or other animal cells (particularly mammalian cells and human cells) such as Vero, MDCK, CV1, 3T3, CPAE, A10, Sp2/0-Ag14, PC12, CHO, COS, HeLa, Bowes melanoma cells, SW-13, NCI-H295, RT4, HT-1376, UM-UC-3, IM-9, KG-1, R54;11, A-172, U-87MG, BT-20, MCF-7, SK-BR-3, ChaGo K-1, CCD-14B, CaSKI, ME-180, FHC, HT-29, Caco-2, SW480, HuTu80, Tera 1, NTERA-2, ANS CA, KLE, RL95-2, Caki-1, ACHN, 769 P, CCRF-CEM, Hut 78, MOLT 4, HL-60, Hep-3B, HepG2, SK-HEP1, A-549, NCI-H146, NCI-H82, NCI-H82, SK-LU-1, WI-38, MRC-5, HLF-a, CCD-19Lh, C39, Ha294T, SK-MEL5, COLO 829, U266B1, RPMI 2650, BeWo, JEG-3, JAR, SW 1533, MeKam, and SCC-4; and higher plant cells. Appropriate culture media and conditions for the above-described host cells are known in the art.

[0078] De Haan et al., *J. Virol.* 72: 6838-50 (1998), describe the assembly of coronavirus VLPs from the coexpression of mouse hepatitis virus M and E genes in eukaryotic cells. Bos et al., *J. Virol.* 71: 9427-33 describe the role of the S protein in infectivity of coronavirus VLPs produced by coexpression of mouse hepatitis virus S, M, and E proteins. These references are hereby incorporated by reference in their entireties.

[0079] In another embodiment, the VLP comprising SARS-CoV polypeptides S, M, and E provides a method for mimicking a SARS-CoV infection without the use of the actual infectious agent. In addition, the VLP provides a method for eliciting a detectable immune response to multiple antigens in a confirmation similar to the actual virus particle thereby enhancing the immunogenicity of the SARS-CoV polypeptides.

[0080] The VLPs of the invention can be produced in vivo by delivery of S, M or E polynucleotides or polypeptides, described herein, to a vertebrate wherein assembly of the VLPs occurs with the cells of the vertebrate. In an alternative embodiment, VLPs of the invention can be produced in vitro in cells that have received the S, M, and E polynucleotides described herein and express said proteins. VLPs are then purified from the cells using techniques known in the art for coronavirus particle purification. These purified particles can then be administered to a vertebrate to elicit a detectable immune response or to study the pathogenesis of the SARS-CoV infection without the need of the actual infectious agent.

[0081] The combination of S, M and E to create virus like particles in the previous examples is not meant to be limiting. Other SARS-CoV polypeptides, which assemble into, or are engineered to assemble into virus like particles, may be used as well.

[0082] The present invention also provides vaccine compositions and methods for delivery of SARS-CoV coding sequences to a vertebrate. In other embodiments, the present

invention provides vaccine compositions and methods for delivery of SARS-CoV coding sequences to a vertebrate with optimal expression and safety conferred through codon optimization and/or other manipulations. These vaccine compositions are prepared and administered in such a manner that the encoded gene products are optimally expressed in the vertebrate of interest. As a result, these compositions and methods are useful in stimulating an immune response against SARS-CoV infection. Also included in the invention are expression systems, delivery systems, and codon-optimized SARS-CoV coding regions.

[0083] In a specific embodiment, the invention provides polynucleotide (e.g., DNA) vaccines in which the single formulation comprises a SARS-CoV polypeptide-encoding polynucleotide vaccine as described herein. An alternative embodiment of the invention provides for a multivalent formulation comprising several (e.g., two, three, four, or more) SARS-CoV polypeptide-encoding polynucleotides, as described herein, within a single vaccine composition. The SARS-CoV polypeptide-encoding polynucleotides, fragments, or variants thereof may be contained within a single expression vector (e.g., plasmid or viral vector) or may be contained within multiple expression vectors.

[0084] In a specific embodiment, the invention provides combinatorial polynucleotide (e.g., DNA) vaccines which combine both a polynucleotide vaccine and polypeptide (e.g., either a recombinant protein, a purified subunit protein, a viral vector expressing an isolated SARS-CoV polypeptide) vaccine in a single formulation. The single formulation comprises a SARS-CoV polypeptide-encoding polynucleotide vaccine as described herein, and optionally, an effective amount of a desired isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof. The polypeptide may exist in any form, for example, a recombinant protein, a purified subunit protein, or a viral vector expressing an isolated SARS-CoV polypeptide. The SARS-CoV polypeptide or fragment, variant, or derivative thereof encoded by the polynucleotide vaccine may be identical to the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof. Alternatively, the SARS-CoV polypeptide or fragment, variant, or derivative thereof encoded by the polynucleotide may be different from the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof.

[0085] It is to be noted that the term "a" or "an" entity refers to one or more of that entity; for example, "a polynucleotide," is understood to represent one or more polynucleotides. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

[0086] It is to be noted that the term "about" when referring to a polynucleotide, coding region or any nucleotide sequence, for example, is understood to represent plus or minus 1 to 30 nucleotides on either end of the defined coding region, polynucleotide or nucleotide sequence. It is to be noted that when referring to a polypeptide, or polypeptide sequence, that the term "about" is understood to represent plus or minus 1 to 10 amino acids on either end of the defined polypeptide or polypeptide sequence. It should be further noted that the term "about," when referring to the quantity of a specific codon in a given codon-optimized coding region has a specific meaning, described in more detail below.

[0087] The term "polynucleotide" is intended to encompass a singular nucleic acid or nucleic acid fragment as well

as plural nucleic acids or nucleic acid fragments, and refers to an isolated molecule or construct, e.g., a virus genome (e.g., a non-infectious viral genome), messenger RNA (mRNA), plasmid DNA (pDNA), or derivatives of pDNA (e.g., minicircles as described in Darquet, A-M et al., *Gene Therapy* 4:1341-1349 (1997)) comprising a polynucleotide. A nucleic acid or fragment thereof may be provided in linear (e.g., mRNA), circular (e.g., plasmid), or branched form as well as double-stranded or single-stranded forms. A polynucleotide may comprise a conventional phosphodiester bond or a non-conventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)).

[0088] The terms "nucleic acid" or "nucleic acid fragment" refer to any one or more nucleic acid segments, e.g., DNA or RNA fragments, present in a polynucleotide or construct.

[0089] As used herein, a "coding region" is a portion of nucleic acid which consists of codons translated into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is not translated into an amino acid, it may be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, and the like, are not part of a coding region. Two or more nucleic acids or nucleic acid fragments of the present invention can be present in a single polynucleotide construct, e.g., on a single plasmid, or in separate polynucleotide constructs, e.g., on separate (different) plasmids. Furthermore, any nucleic acid or nucleic acid fragment may encode a single SARS-CoV polypeptide or fragment, derivative, or variant thereof, e.g., or may encode more than one polypeptide, e.g., a nucleic acid may encode two or more polypeptides. In addition, a nucleic acid may include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator, or may encode heterologous coding regions fused to the SARS-CoV coding region, e.g., specialized elements or motifs, such as a secretory signal peptide or a heterologous functional domain.

[0090] The terms "fragment," "variant," "derivative," and "analog," when referring to SARS-CoV polypeptides of the present invention, include any polypeptides which retain at least some of the immunogenicity or antigenicity of the corresponding native polypeptide. Fragments of SARS-CoV polypeptides of the present invention include proteolytic fragments, deletion fragments, and in particular, fragments of SARS-CoV polypeptides which exhibit increased secretion from the cell or higher immunogenicity or reduced pathogenicity when delivered to an animal. Polypeptide fragments further include any portion of the polypeptide which comprises an antigenic or immunogenic epitope of the native polypeptide, including linear as well as three-dimensional epitopes. Variants of SARS-CoV polypeptides of the present invention include fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants may occur naturally, such as an allelic variant. By an "allelic variant" is intended alternate forms of a gene occupying a given locus on a chromosome or genome of an organism or virus. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985), which is incorporated herein by reference. Naturally or non-naturally occurring variations such as amino acid deletions, insertions or substitutions may occur. Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Variant polypep-

tides may comprise conservative or non-conservative amino acid substitutions, deletions or additions. Derivatives of SARS-CoV polypeptides of the present invention, are polypeptides which have been altered so as to exhibit additional features not found on the native polypeptide. Examples include fusion proteins. An analog is another form of a SARS-CoV polypeptide of the present invention. An example is a proprotein which can be activated by cleavage of the proprotein to produce an active mature polypeptide.

[0091] The terms "infectious polynucleotide" or "infectious nucleic acid" are intended to encompass isolated viral polynucleotides and/or nucleic acids which are solely sufficient to mediate the synthesis of complete infectious virus particles upon uptake by permissive cells. Thus, "infectious nucleic acids" do not require pre-synthesized copies of any of the polypeptides it encodes, e.g., viral replicases, in order to initiate its replication cycle in a permissive host cell.

[0092] The terms "non-infectious polynucleotide" or "non-infectious nucleic acid" as defined herein are polynucleotides or nucleic acids which cannot, without additional added materials, e.g., polypeptides, mediate the synthesis of complete infectious virus particles upon uptake by permissive cells. An infectious polynucleotide or nucleic acid is not made "non-infectious" simply because it is taken up by a non-permissive cell. For example, an infectious viral polynucleotide from a virus with limited host range is infectious if it is capable of mediating the synthesis of complete infectious virus particles when taken up by cells derived from a permissive host (i.e., a host permissive for the virus itself). The fact that uptake by cells derived from a non-permissive host does not result in the synthesis of complete infectious virus particles does not make the nucleic acid "non-infectious." In other words, the term is not qualified by the nature of the host cell, the tissue type, or the species taking up the polynucleotide or nucleic acid fragment.

[0093] In some cases, an isolated infectious polynucleotide or nucleic acid may produce fully-infectious virus particles in a host cell population which lacks receptors for the virus particles, i.e., is non-permissive for virus entry.

[0094] Thus viruses produced will not infect surrounding cells. However, if the supernatant containing the virus particles is transferred to cells which are permissive for the virus, infection will take place.

[0095] The terms "replicating polynucleotide" or "replicating nucleic acid" are meant to encompass those polynucleotides and/or nucleic acids which, upon being taken up by a permissive host cell, are capable of producing multiple, e.g., one or more copies of the same polynucleotide or nucleic acid. Infectious polynucleotides and nucleic acids are a subset of replicating polynucleotides and nucleic acids; the terms are not synonymous. For example, a defective virus genome lacking the genes for virus coat proteins may replicate, e.g., produce multiple copies of itself, but is NOT infectious because it is incapable of mediating the synthesis of complete infectious virus particles unless the coat proteins, or another nucleic acid encoding the coat proteins, are exogenously provided.

[0096] In certain embodiments, the polynucleotide, nucleic acid, or nucleic acid fragment is a nucleic acid which DNA, a polynucleotide comprising a nucleic acid which

encodes a polypeptide normally also comprises a promoter and/or other transcription or translation control elements operably associated with the polypeptide-encoding nucleic acid fragment. An operable association is when a nucleic acid fragment encoding a gene product, e.g., a polypeptide, is associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide-encoding nucleic acid fragment and a promoter associated with the 5' end of the nucleic acid fragment) are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the expression regulatory sequences to direct the expression of the gene product, or (3) interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a nucleic acid fragment encoding a polypeptide if the promoter were capable of effecting transcription of that nucleic acid fragment. The promoter may be a cell-specific promoter that directs substantial transcription of the DNA only in predetermined cells. Other transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide to direct cell-specific transcription. Suitable promoters and other transcription control regions are disclosed herein.

[0097] A variety of transcription control regions are known to those skilled in the art. These include, without limitation, transcription control regions which function in vertebrate cells, such as, but not limited to, promoter and enhancer segments from cytomegaloviruses (the immediate early promoter, in conjunction with intron-A), simian virus 40 (the early promoter), and retroviruses (such as Rous sarcoma virus). Other transcription control regions include those derived from vertebrate genes such as actin, heat shock protein, bovine growth hormone and rabbit β -globin, as well as other sequences capable of controlling gene expression in eukaryotic cells. Additional suitable transcription control regions include tissue-specific promoters and enhancers as well as lymphokine-inducible promoters (e.g., promoters inducible by interferons or interleukins).

[0098] Similarly, a variety of translation control elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, elements from picornaviruses (particularly an internal ribosome entry site, or IRES, also referred to as a CITE sequence).

[0099] A DNA polynucleotide of the present invention may be a circular or linearized plasmid, or other linear DNA which may also be non-infectious and nonintegrating (i.e., does not integrate into the genome of vertebrate cells). A linearized plasmid is a plasmid that was previously circular but has been linearized, for example, by digestion with a restriction endonuclease. Linear DNA may be advantageous in certain situations as discussed, e.g., in Cherng, J. Y., et al., *J. Control. Release* 60:343-53 (1999), and Chen, Z. Y., et al., *Mol. Ther.* 3:403-10 (2001), both of which are incorporated herein by reference.

[0100] Alternatively, DNA virus genomes may be used to administer DNA polynucleotides into vertebrate cells. In

certain embodiments, a DNA virus genome of the present invention is nonreplicative, noninfectious, and/or nonintegrating. Suitable DNA virus genomes include without limitation, herpesvirus genomes, adenovirus genomes, adeno-associated virus genomes, and poxvirus genomes. References citing methods for the *in vivo* introduction of non-infectious virus genomes to vertebrate tissues are well known to those of ordinary skill in the art, and are cited *supra*.

[0101] In other embodiments, a polynucleotide of the present invention is RNA, for example, in the form of messenger RNA (mRNA). Methods for introducing RNA sequences into vertebrate cells are described in U.S. Pat. No. 5,580,859, the disclosure of which is incorporated herein by reference in its entirety.

[0102] Polynucleotides, nucleic acids, and nucleic acid fragments of the present invention may be associated with additional nucleic acids which encode secretory or signal peptides, which direct the secretion of a polypeptide encoded by a nucleic acid fragment or polynucleotide of the present invention. According to the signal hypothesis, proteins secreted by mammalian cells have a signal peptide or secretory leader sequence which is cleaved from the mature protein once export of the growing protein chain across the rough endoplasmic reticulum has been initiated. Those of ordinary skill in the art are aware that polypeptides secreted by vertebrate cells generally have a signal peptide fused to the N-terminus of the polypeptide, which is cleaved from the complete or "full length" polypeptide to produce a secreted or "mature" form of the polypeptide. In certain embodiments, the native leader sequence is used, or a functional derivative of that sequence that retains the ability to direct the secretion of the polypeptide that is operably associated with it. Alternatively, a heterologous mammalian leader sequence, or a functional derivative thereof, may be used. For example, the wild-type leader sequence may be substituted with the leader sequence of human tissue plasminogen activator (TPA) or mouse β -glucuronidase.

[0103] In accordance with one aspect of the present invention, there is provided a polynucleotide construct, for example, a plasmid, comprising a nucleic acid fragment, where the nucleic acid fragment is a fragment of a coding region operably encoding an SARS-CoV-derived polypeptide. In accordance with another aspect of the present invention, there is provided a polynucleotide construct, for example, a plasmid, comprising a nucleic acid fragment, where the nucleic acid fragment is a fragment of a codon-optimized coding region operably encoding an SARS-CoV-derived polypeptide, where the coding region is optimized for expression in vertebrate cells, of a desired vertebrate species, e.g., humans, to be delivered to a vertebrate to be treated or immunized. Suitable SARS-CoV polypeptides, or fragments, variants, or derivatives thereof may be derived from, but are not limited to, the SARS-CoV S, Soluble S1, Soluble S2, N, E or M proteins. Additional SARS-CoV-derived coding sequences, e.g., coding for S, Soluble S1, Soluble S2, N, E or M, may also be included on the plasmid, or on a separate plasmid, and expressed, either using native SARS-CoV codons or one or more codons optimized for expression in the vertebrate to be treated or immunized. When such a plasmid encoding one or more optimized SARS-CoV sequences and/or one or more optimized SARS-CoV sequences is delivered, *in vivo* to a tissue of the

vertebrate to be treated or immunized, one or more of the encoded gene products will be expressed, i.e., transcribed and translated. The level of expression of the gene product(s) will depend to a significant extent on the strength of the associated promoter and the presence and activation of an associated enhancer element, as well as the degree of optimization of the coding region.

[0104] As used herein, the term "plasmid" refers to a construct made up of genetic material (i.e., nucleic acids). Typically a plasmid contains an origin of replication which is functional in bacterial host cells, e.g., *Escherichia coli*, and selectable markers for detecting bacterial host cells comprising the plasmid. Plasmids of the present invention may include genetic elements as described herein arranged such that an inserted coding sequence can be transcribed and translated in eukaryotic cells. Also, the plasmid may include a sequence from a viral nucleic acid. However, such viral sequences normally are not sufficient to direct or allow the incorporation of the plasmid into a viral particle, and the plasmid is therefore a non-viral vector. In certain embodiments described herein, a plasmid is a closed circular DNA molecule.

[0105] The term "expression" refers to the biological production of a product encoded by a coding sequence. In most cases a DNA sequence, including the coding sequence, is transcribed to form a messenger-RNA (mRNA). The messenger-RNA is then translated to form a polypeptide product which has a relevant biological activity. Also, the process of expression may involve further processing steps to the RNA product of transcription, such as splicing to remove introns, and/or post-translational processing of a polypeptide product.

[0106] As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and comprises any chain or chains of two or more amino acids. Thus, as used herein, terms including, but not limited to "peptide," "dipeptide," "tripeptide," "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids, are included in the definition of a "polypeptide," and the term "polypeptide" may be used instead of, or interchangeably with any of these terms. The term further includes polypeptides which have undergone post-translational modifications, for example, glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-naturally occurring amino acids.

[0107] Also included as polypeptides of the present invention are fragments, derivatives, analogs, or variants of the foregoing polypeptides, and any combination thereof. Polypeptides, and fragments, derivatives, analogs, or variants thereof of the present invention can be antigenic and immunogenic polypeptides related to SARS-CoV polypeptides, which are used to prevent or treat, i.e., cure, ameliorate, lessen the severity of, or prevent or reduce contagion of infectious disease caused by the SARS-CoV.

[0108] As used herein, an antigenic polypeptide or an immunogenic polypeptide is a polypeptide which, when introduced into a vertebrate, reacts with the vertebrate's immune system molecules, i.e., is antigenic, and/or induces an immune response in the vertebrate, i.e., is immunogenic. It is quite likely that an immunogenic polypeptide will also

be antigenic, but an antigenic polypeptide, because of its size or conformation, may not necessarily be immunogenic. Examples of antigenic and immunogenic polypeptides of the present invention include, but are not limited to, e.g., S or fragments, derivatives, or variants thereof; N or fragments, derivatives, or variants thereof; E or fragments, derivatives, or variants thereof; M or fragments, derivatives, or variants thereof; other predicted ORF's within the sequence of the SARS-CoV viruses which may possess antigenic properties, for example, an ORF which may encode for the hemagglutinin-esterase or fragments, derivatives, or variants thereof; or any of the foregoing polypeptides or fragments, derivatives, or variants thereof fused to a heterologous polypeptide, for example, a hepatitis B core antigen. Isolated antigenic and immunogenic polypeptides of the present invention in addition to those encoded by polynucleotides of the invention, may be provided as a recombinant protein, a purified subunit, a viral vector expressing the protein, or may be provided in the form of an inactivated SARS-CoV vaccine, e.g., a live-attenuated virus vaccine, a heat-killed virus vaccine, etc.

[0109] By an "isolated" SARS-CoV polypeptide or a fragment, variant, or derivative thereof is intended a SARS-CoV polypeptide or protein that is not in its natural environment. No particular level of purification is required. For example, an isolated SARS-CoV polypeptide can be removed from its native or natural environment. Recombinantly produced SARS-CoV polypeptides and proteins expressed in host cells are considered isolated for purposes of the invention, as are native or recombinant SARS-CoV polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique, including the separation of SARS-CoV virions from tissue samples or culture cells in which they have been propagated. In addition, an isolated, thus, isolated SARS-CoV polypeptides and proteins can be provided as, for example, recombinant SARS-CoV polypeptides, a purified subunit of SARS-CoV, or a viral vector expressing an isolated SARS-CoV polypeptide.

[0110] The term "epitopes," as used herein, refers to portions of a polypeptide having antigenic or immunogenic activity in a vertebrate, for example a human. An "immunogenic epitope," as used herein, is defined as a portion of a protein that elicits an immune response in an animal, as determined by any method known in the art. The term "antigenic epitope," as used herein, is defined as a portion of a protein to which an antibody or T-cell receptor can immunospecifically bind as determined by any method well known in the art. Immunospecific binding excludes non-specific binding but does not exclude cross-reactivity with other antigens. Where all immunogenic epitopes are antigenic, antigenic epitopes need not be immunogenic.

[0111] The term "immunogenic carrier" as used herein refers to a first polypeptide or fragment, variant, or derivative thereof which enhances the immunogenicity of a second polypeptide or fragment, variant, or derivative thereof. Typically, an "immunogenic carrier" is fused to or conjugated to the desired polypeptide or fragment thereof. An example of an "immunogenic carrier" is a recombinant hepatitis B core antigen expressing, as a surface epitope, an immunogenic epitope of interest. See, e.g., European Patent No. EP 0385610 B 1, which is incorporated herein by reference in its entirety.

[0112] In the present invention, antigenic epitopes preferably contain a sequence of at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, or between about 8 to about 30 amino acids contained within the amino acid sequence of a SARS-CoV polypeptide of the invention, e.g., an S polypeptide, an N polypeptide, an E polypeptide or an M polypeptide. Certain polypeptides comprising immunogenic or antigenic epitopes are at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acid residues in length. Antigenic as well as immunogenic epitopes may be linear, i.e., be comprised of contiguous amino acids in a polypeptide, or may be three dimensional, i.e., where an epitope is comprised of non-contiguous amino acids which come together due to the secondary or tertiary structure of the polypeptide, thereby forming an epitope.

[0113] As to the selection of peptides or polypeptides bearing an antigenic epitope (e.g., that contain a region of a protein molecule to which an antibody or T cell receptor can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, e.g., Sutcliffe, J. G., et al., *Science* 219:660-666 (1983).

[0114] Peptides capable of eliciting an immunogenic response are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the mimicked protein; longer peptides, especially those containing proline residues, usually are effective. Sutcliffe et al., supra, at 661. For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HA1 polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the rabies glycoprotein induced antibodies that precipitated the respective proteins.

Codon Optimization

[0115] "Codon optimization" is defined as modifying a nucleic acid sequence for enhanced expression in the cells of the vertebrate of interest, e.g., human, by replacing at least one, more than one, or a significant number, of codons of the native sequence with codons that are more frequently or most frequently used in the genes of that vertebrate. Various species exhibit particular biases for certain codons of a particular amino acid.

[0116] In one aspect, the present invention relates to polynucleotides comprising nucleic acid fragments of codon-optimized coding regions which encode SARS-CoV polypeptides, or fragments, variants, or derivatives thereof, with the codon usage adapted for optimized expression in the cells of a given vertebrate, e.g., humans. These polynucleotides are prepared by incorporating codons preferred for use in the genes of the vertebrate of interest into the DNA sequence. Also provided are polynucleotide expression constructs, vectors, and host cells comprising nucleic acid fragments of codon-optimized coding regions which encode SARS-CoV polypeptides, and fragments, variants, or

derivatives thereof, and various methods of using the polynucleotide expression constructs, vectors, and/or host cells to treat or prevent SARS disease in a vertebrate.

[0117] As used herein the term "codon-optimized coding region" means a nucleic acid coding region that has been adapted for expression in the cells of a given vertebrate by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that vertebrate.

[0118] Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code," which shows which codons encode which amino acids, is reproduced herein as Table 3. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six triplets, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

TABLE 3

The Standard Genetic Code				
T	C	A	G	
T	TTT Phe (F) TCT Ser (S) TAT Tyr (Y) TGT Cys (C)	TTC Phe (F) TCC Ser (S) TAC Tyr (Y) TGC	TTA Leu (L) TCA Ser (S) TAA Ter	TGA Ter
	TTG Leu (L) TCG Ser (S) TAG Ter		TGG Trp (W)	
C	CTT Leu (L) CCT Pro (P) CAT His (H) COT Arg (R)	CTC Leu (L) CCC Pro (P) CAC His (H) CCG Arg (R)	CTA Leu (L) CCA Pro (P) CAA Gln (Q) CGA Arg (R)	CTG Leu (L) CCG Pro (P) CAG Gln (Q) CGG Arg (R)
A	ATT Ile (I) ACT Thr (T) AAT Asn (N) AOT Ser (S)	ATC Ile (I) ACC Thr (T) AAC Asn (N) AGC Ser (S)	ATA Ile (I) ACA Thr (T) AAA Lys (K) AGA Arg (R)	ATG Met (M) ACG Thr (T) AAG Lys (K) AAG Arg (R)
G	GTT Val (V) GCT Ala (A) GAT Asp (D) GGT Gly (G)	GTC Val (V) GGC Ala (A) GAC Asp (D) GGG Gly (G)	GTA Val (V) GCA Ala (A) GAA Glu (E) GGA Gly (G)	GTG Val (V) GCG Ala (A) GAG Glu (E) GGG Gly (G)

[0119] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0120] Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative frequencies of codon usage. Codon usage tables are readily available, for example, at the "Codon Usage Database," available at <http://www.kazusa.or.jp/codon/> (visited Jul. 9, 2002), and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). As examples, the codon usage tables for human, mouse, domestic cat, and cow, calculated from GenBank Release 128.0 (15 Feb. 2002), are reproduced below as Tables 4-7. These tables use mRNA nomenclature, and so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The tables have been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

TABLE 4

<u>Codon Usage Table for Human Genes (<i>Homo sapiens</i>)</u>			
Amino Acid	Codon	Number	Frequency
Phe	UUU	326146	0.4525
Phe	UUC	394680	0.5475
Total		720826	
Leu	UUA	139249	0.0728
Leu	UUG	242151	0.1286
Leu	CUU	246206	0.1287
Leu	CUC	374262	0.1956
Leu	CUA	133980	0.0700
Leu	CUG	777077	0.4062
Total		1912925	
Ile	AUU	303721	0.3554
Ile	AUC	414483	0.4850
Ile	AUA	136399	0.1596
Total		854603	
Met	AUG	430946	1.0000
Total		430946	
Val	GUU	210423	0.1773
Val	GUC	282445	0.2380
Val	GUA	134991	0.1137
Val	GUG	559044	0.4710
Total		1186903	
Ser	UCU	282407	0.1840
Ser	UCC	336349	0.2191
Ser	UCA	225963	0.1472
Ser	UCG	86761	0.0565
Ser	AGU	230047	0.1499
Ser	AGC	373362	0.2433
Total		1534889	
Pro	CCU	333705	0.2834
Pro	CCC	386462	0.3281
Pro	CCA	322220	0.2736
Pro	CCG	135317	0.1149
Total		1177704	
Thr	ACU	247913	0.2419
Thr	ACC	371420	0.3624
Thr	ACA	285535	0.2787
Thr	ACG	120022	0.1171
Total		1025010	
Ala	GCU	360146	0.2637
Ala	GCC	551432	0.40370

TABLE 4-continued

<u>Codon Usage Table for Human Genes (<i>Homo sapiens</i>)</u>			
Amino Acid	Codon	Number	Frequency
Ala	GCA	308014	0.2255
Ala	GCG	146133	0.1071
Total		1365865	
Tyr	UAU	232240	0.4347
Tyr	UAC	301978	0.5653
Total		534218	
His	CAU	201389	0.4113
His	CAC	288200	0.5887
Total		489589	
Gln	CAA	227742	0.2541
Gln	CAG	668391	0.7459
Total		896133	
Asn	AAU	322271	0.4614
Asn	AAC	376210	0.5386
Total		698481	
Lys	AAA	462660	0.4212
Lys	AAG	635755	0.5788
Total		1098415	
Asp	GAU	430744	0.4613
Asp	GAC	502940	0.5387
Total		933684	
Glu	GAA	651277	0.4161
Glu	GAG	787712	0.5839
Total		1348989	
Cys	UGU	190952	0.4468
Cys	UGC	236400	0.5532
Total		427352	
Trp	UGG	248083	1.0000
Total		248083	
Arg	CGU	90899	0.0830
Arg	CGC	210931	0.1927
Arg	CGA	122355	0.1120
Arg	CGG	228970	0.2092
Arg	AGA	221221	0.2021
Arg	AGG	220119	0.2011
Total		1094695	
Gly	GGU	209450	0.1632
Gly	GGC	441320	0.3438
Gly	OGA	315726	0.2459
Gly	GGG	317263	0.2471
Total		1283759	
Stop	UAA	13963	
Stop	UAG	10631	
Stop	UGA	24607	

[0121]

TABLE 5

<u>Codon Usage Table for Mouse Genes (<i>Mus musculus</i>)</u>			
Amino Acid	Codon	Number	Frequency
Phe	UUU	150467	0.4321
Phe	UUC	197795	0.5679
Total		348262	

TABLE 5-continued

Codon Usage Table for Mouse Genes (<i>Mus musculus</i>)			
Amino Acid	Codon	Number	Frequency
Leu	UUA	55635	0.0625
Leu	UUG	116210	0.1306
Leu	CUU	114699	0.1289
Leu	CUC	179248	0.2015
Leu	CUA	69237	0.0778
Leu	CUG	354743	0.3987
Total		889772	
Ile	AUU	137513	0.3367
Ile	AUC	208533	0.5106
Ile	AUA	62349	0.1527
Total		408395	
Met	AUG	204546	1.0000
Total		204546	
Val	GUU	93754	0.1673
Val	GUC	140762	0.2513
Val	GUA	64417	0.1150
Val	GUG	261308	0.4664
Total		560241	
Ser	UCU	139576	0.1936
Ser	UCC	160313	0.2224
Ser	UCA	109524	0.1394
Ser	UCG	38632	0.0536
Ser	AGU	108413	0.1504
Ser	AGC	173518	0.2407
Total		720976	
Pro	CCU	162613	0.3036
Pro	CCC	164796	0.3077
Pro	CCA	151091	0.2821
Pro	CCG	57032	0.1065
Total		535532	
Thr	ACU	119832	0.2472
Thr	ACC	172415	0.3556
Thr	ACA	140420	0.2896
Thr	ACG	52142	0.1076
Total		484809	
Ala	GCU	178593	0.2905
Ala	GCC	236018	0.3839
Ala	GCA	139697	0.2272
Ala	GCG	60444	0.0983
Total		614752	
Tyr	UAU	108556	0.4219
Tyr	UAC	148772	0.5781
Total		257328	
His	CAU	88786	0.3973
His	CAC	134705	0.6027
Total		223491	
Gln	CAA	101783	0.2520
Gln	CAG	302064	0.7480
Total		403847	
Asn	AAU	138868	0.4254
Asn	AAC	187541	0.5746
Total		326409	
Lys	AAA	188707	0.3839
Lys	AAG	302799	0.6161
Total		491506	
Asp	GAU	189172	0.4414
Asp	GAC	239670	0.5586
Total		429042	

TABLE 5-continued

Codon Usage Table for Mouse Genes (<i>Mus musculus</i>)			
Amino Acid	Codon	Number	Frequency
Glu	GAA	235842	0.4015
Glu	GAG	351582	0.5985
Total		587424	
Cys	UGU	97385	0.4716
Cys	UGC	109130	0.5284
Total		206515	
Trp	UGG	112588	1.0000
Total		112588	
Arg	CGU	41703	0.0863
Arg	CGC	86351	0.1787
Arg	CGA	38928	0.1220
Arg	CGG	92277	0.1910
Arg	AGA	101029	0.2091
Arg	AGG	102859	0.2129
Total		483147	
Gly	GGU	103673	0.1750
Gly	GGC	196604	0.3352
Gly	GGA	151497	0.2557
Gly	GGG	138700	0.2341
Total		592474	
Stop	UAA	5499	
Stop	UAG	4661	
Stop	UGA	10356	

[0122]

TABLE 6

Codon Usage Table for Domestic Cat Genes (<i>Felis catus</i>)			
Amino Acid	Codon	Number	Frequency of usage
Phe	UUU	1204.00	0.4039
Phe	UUC	1777.00	0.5961
Total		2981	
Leu	UUA	404.00	0.0570
Leu	UUG	857.00	0.1209
Leu	CUU	791.00	0.1116
Leu	CUC	1513.00	0.2135
Leu	CUA	488.00	0.0688
Leu	CUG	3035.00	0.4282
Total		7088	
Ile	AUU	1018.00	0.2984
Ile	AUC	1815.00	0.5380
Ile	AUA	558.00	0.1636
Total		3411	
Met	AUG	1553.00	0.0636
Total		1553	
Val	GUU	696.00	0.1512
Val	GUC	1279.00	0.2779
Val	GUA	463.00	0.1026
Val	GUG	2164.00	0.4702
Total		4602	
Ser	UCU	940.00	0.1875
Ser	UCC	1260.00	0.2513
Ser	UCA	608.00	0.1213
Ser	UCG	332.00	0.0662

TABLE 6-continued

Codon Usage Table for Domestic Cat Genes (<i>Felis catus</i>)			
Amino Acid	Codon	Number	Frequency of usage
Ser	AGU	672.00	0.1340
Ser	AGC	1202.00	0.2397
Total		5014	
Pro	CCU	958.00	0.2626
Pro	CCC	1375.00	0.3769
Pro	CCA	850.00	0.2330
Pro	CCG	465.00	0.1275
Total		3648	
Thr	ACU	822.00	0.2127
Thr	ACC	1574.00	0.4072
Thr	ACA	903.00	0.2336
Thr	ACG	566.00	0.1464
Total		3865	
Ala	GCU	1129.00	0.2496
Ala	GCC	1951.00	0.4313
Ala	GCA	883.00	0.1952
Ala	GCG	561.00	0.1240
Total		4524	
Tyr	UAU	837.00	0.3779
Tyr	UAC	1378.00	0.6221
Total		2215	
His	CAU	594.00	0.3738
His	CAC	995.00	0.6262
Total		1589	
Gln	CAA	747.00	0.2783
Gln	CAG	1937.00	0.7217
Total		2684	
Asn	AAU	1109.00	0.3949
Asn	AAC	1699.00	0.6051
Total		2808	
Lys	AAA	1445.00	0.4088
Lys	AAG	2090.00	0.5912
Total		3535	
Asp	GAU	1255.00	0.4055
Asp	GAC	1840.00	0.5945
Total		3095	
Glu	GAA	1637.00	0.4164
Glu	GAG	2294.00	0.5836
Total		3931	
Cys	UGU	719.00	0.4425
Cys	UGC	906.00	0.5575
Total		1625	
Trp	UGG	1073.00	1.0000
Total		1073	
Arg	CGU	236.00	0.0700
Arg	CGC	629.00	0.1865
Arg	CGA	354.00	0.1050
Arg	CGG	662.00	0.1963
Arg	AGA	712.00	0.2112
Arg	AGG	779.00	0.2310
Total		3372	
Gly	GGU	648.00	0.1498
Gly	GGC	1536.00	0.3551
Gly	GGA	1065.00	0.2462
Gly	GGG	1077.00	0.2490
Total		4326	

TABLE 6-continued

Codon Usage Table for Domestic Cat Genes (<i>Felis catus</i>)			
Amino Acid	Codon	Number	Frequency of usage
Stop	UAA	55	
Stop	UAG	36	
Stop	UGA	110	

[0123]

TABLE 7

Codon Usage Table for Cow Genes (<i>Bos taurus</i>)			
Amino Acid	Codon	Number	Frequency of usage
Phe	UUU	13002	0.4112
Phe	UUC	18614	0.5888
Total		31616	
Leu	UUA	4467	0.0590
Leu	UUG	9024	0.1192
Leu	CUU	9069	0.1198
Leu	CUC	16003	0.2114
Leu	CUA	4658	0.0609
Leu	CUG	32536	0.4298
Total		75707	
Ile	AUU	12474	0.3313
Ile	AUC	19800	0.5258
Ile	AUA	5381	0.1429
Total		37655	
Met	AUG	17770	1.0000
Total		17770	
Val	GUU	8212	0.1635
Val	GUC	12846	0.2558
Val	GUA	4932	0.0982
Val	GUG	24222	0.4824
Total		50212	
Ser	UCU	10287	0.1804
Ser	UCC	13258	0.2325
Ser	UCA	7678	0.1347
Ser	UCG	3470	0.0609
Ser	AGU	8040	0.1410
Ser	AGC	14279	0.2505
Total		57012	
Pro	CCU	11695	0.2684
Pro	CCC	15221	0.3493
Pro	CCA	11039	0.2533
Pro	CCG	5621	0.1290
Total		43576	
Thr	ACU	9372	0.2203
Thr	ACC	16574	0.3895
Thr	ACA	10892	0.2560
Thr	ACG	5712	0.1342
Total		42550	
Ala	GCU	13923	0.2592
Ala	GCC	23073	0.4295
Ala	GCA	10704	0.1992
Ala	GCG	6025	0.1121
Total		53725	
Tyr	UAU	9441	0.3882
Tyr	UAC	14882	0.6118
Total		24323	

TABLE 7-continued

Codon Usage Table for Cow Genes (<i>Bos taurus</i>)			
Amino Acid	Codon	Number	Frequency of usage
His	CAU	6528	0.3649
His	CAC	11363	0.6351
Total		17891	
Gln	CAA	8060	0.2430
Gln	CAG	25108	0.7570
Total		33168	
Asn	AAU	12491	0.4088
Asn	AAC	18063	0.5912
Total		30554	
Lys	AAA	17244	0.3897
Lys	AAG	27000	0.6103
Total		44244	
Asp	GAU	16615	0.4239
Asp	GAC	22380	0.5761
Total		39195	
Glu	GAA	21102	0.4007
Glu	GAG	31553	0.5993
Total		52657	
Cys	UGU	7556	0.4200
Cys	UGC	10436	0.5800
Total		17992	
Trp	UGG	10706	1.0000
Total		10706	
Arg	CGU	3391	0.0824
Arg	CGC	7998	0.1943
Arg	CGA	4558	0.1108
Arg	CGG	8300	0.2017
Arg	AGA	8237	0.2001
Arg	AGG	8671	0.2107
Total		41155	
Gly	GGU	8508	0.1616
Gly	GGC	18517	0.3518
Gly	GGA	12838	0.2439
Gly	GGG	12772	0.2427
Total		52635	
Stop	UAA	555	
Stop	UAG	394	
Stop	UGA	392	

[0124] By utilizing these or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide, but which uses codons more optimal for a given species. Codon-optimized coding regions can be designed by various different methods.

[0125] In one method, termed "uniform optimization," a codon usage table is used to find the single most frequent codon used for any given amino acid, and that codon is used each time that particular amino acid appears in the polypeptide sequence. For example, referring to Table 4 above, the most frequent codon for leucine in humans is CUG, which is used 41% of the time. Thus, all of the leucine residues in a given amino acid sequence would be assigned the codon CUG. A coding region for SARS-CoV soluble S protein (SEQ ID NO:1) optimized by the "uniform optimization" method is presented herein as SEQ ID NO:25.

[0126] In another method, termed "full-optimization," the actual frequencies of the codons are distributed randomly throughout the coding region. Thus, using this method for optimization, if a hypothetical polypeptide sequence had 100 leucine residues, referring to Table 4 for frequency of usage in humans, about 7, or 7% of the leucine codons would be UUA, about 13, or 13% of the leucine codons would be UUG, about 13, or 13% of the leucine codons would be CUU, about 20, or 20% of the leucine codons would be CUC, about 7, or 7% of the leucine codons would be CUA, and about 41, or 41% of the leucine codons would be CUG. These frequencies would be distributed randomly throughout the leucine codons in the coding region encoding the hypothetical polypeptide. As will be understood by those of ordinary skill in the art, the distribution of codons in the sequence can vary significantly using this method, however, the sequence always encodes the same polypeptide.

[0127] As an example, a nucleotide sequence for soluble S (SEQ ID NO:1) fully optimized for human codon usage, is shown as SEQ ID NO:24.

[0128] In using the "full-optimization" method, an entire polypeptide sequence may be codon-optimized as described above. With respect to various desired fragments, variants, or derivatives of the complete polypeptide, the fragment, variant, or derivative may first be designed, and is then codon-optimized individually. Alternatively, a full-length polypeptide sequence is codon-optimized for a given species, resulting in a codon-optimized coding region encoding the entire polypeptide; then nucleic acid fragments of the codon-optimized coding region, which encode fragments, variants, and derivatives of the polypeptide, are made from the original codon-optimized coding region. As will be well understood by those of ordinary skill in the art, if codons have been randomly assigned to the full-length coding region based on their frequency of use in a given species, nucleic acid fragments encoding fragments, variants, and derivatives would not necessarily be fully codon-optimized for the given species. However, such sequences are still much closer to the codon usage of the desired species than the native codon usage. The advantage of this approach is that synthesizing codon-optimized nucleic acid fragments encoding each fragment, variant, and derivative of a given polypeptide, although routine, would be time consuming and would result in significant expense.

[0129] When using the "full-optimization" method, the term "about" is used precisely to account for fractional percentages of codon frequencies for a given amino acid. As used herein, "about" is defined as one amino acid more or one amino acid less than the value given. The whole number value of amino acids is rounded up if the fractional frequency of usage is 0.50 or greater, and is rounded down if the fractional frequency of use is 0.49 or less. Using again the example of the frequency of usage of leucine in human genes, for a hypothetical polypeptide having 62 leucine residues, the fractional frequency of codon usage would be calculated by multiplying 62 by the frequencies for the various codons. Thus, 7.28 percent of 62 equals 4.51 UUA codons, or "about 5," i.e., 4, 5, or 6 UUA codons, 12.66 percent of 62 equals 7.85 UUG codons or "about 8," i.e., 7, 8, or 9 UUG codons, 12.87 percent of 62 equals 7.98 CUU codons, or "about 8," i.e., 7, 8, or 9 CUU codons, 19.56 percent of 62 equals 12.13 CUC codons or "about 12," i.e., 11, 12, or 13 CUC codons, 7.00 percent of 62 equals 4.34

CUA codons or "about 4," i.e., 3, 4, or 5 CUA codons, and 40.62 percent of 62 equals 25.19 CUG codons, or "about 25," i.e., 24, 25, or 26 CUG codons.

[0130] In a third method termed "minimal optimization," coding regions are only partially optimized. For example, the invention includes a nucleic acid fragment of a codon-optimized coding region encoding a polypeptide in which at least about 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the codon positions have been codon-optimized for a given species. That is, they contain a codon that is preferentially used in the genes of a desired species, e.g., a vertebrate species, e.g., humans, in place of a codon that is normally used in the native nucleic acid sequence. Codons that are rarely found in the genes of the vertebrate of interest are changed to codons more commonly utilized in the coding regions of the vertebrate of interest.

[0131] Thus, those codons which are used more frequently in the SARS-CoV gene of interest than in genes of the vertebrate of interest are substituted with more frequently-used codons. The difference in frequency at which the SARS-CoV codons are substituted may vary based on a number factors as discussed below. For example, codons used at least twice more per thousand in SARS-CoV genes as compared to genes of the vertebrate of interest are substituted with the most frequently used codon for that amino acid in the vertebrate of interest. This ratio may be adjusted higher or lower depending on various factors such as those discussed below. Accordingly, a codon in a SARS-CoV native coding region would be substituted with a codon used more frequently for that amino acid in coding regions of the vertebrate of interest if the codon is used 1.1 times, 1.2 times, 1.3 times, 1.4 times, 1.5 times, 1.6 times, 1.7 times, 1.8 times, 1.9 times, 2.0 times, 2.1 times, 2.2 times, 2.3 times, 2.4 times, 2.5 times, 2.6 times, 2.7 times, 2.8 times, 2.9 times, 3.0 times, 3.1 times, 3.2 times, 3.3 times, 3.4 times, 3.5 times, 3.6 times, 3.7 times, 3.8 times, 3.9 times, 4.0 times, 4.1 times, 4.2 times, 4.3 times, 4.4 times, 4.5 times, 4.6 times, 4.7 times, 4.8 times, 4.9 times, 5.0 times, 5.5 times, 6.0 times, 6.5 times, 7.0 times, 7.5 times, 8.0 times, 8.5 times, 9.0 times, 9.5 times, 10.0 times, 10.5 times, 11.0 times, 11.5 times, 12.0 times, 12.5 times, 13.0 times, 13.5 times, 14.0 times, 14.5 times, 15.0 times, 15.5 times, 16.0 times, 16.5 times, 17.0 times, 17.5 times, 18.0 times, 18.5 times, 19.0 times, 19.5 times, 20 times, 21 times, 22 times, 23 times, 24 times, 25 times, or greater more frequently in SARS-CoV coding regions than in coding regions of the vertebrate of interest.

[0132] This minimal human codon optimization for highly variant codons has several advantages, which include but are not limited to the following examples. Since fewer changes are made to the nucleotide sequence of the gene of interest, fewer manipulations are required, which leads to reduced risk of introducing unwanted mutations and lower cost, as well as allowing the use of commercially available site-directed mutagenesis kits, and reducing the need for expensive oligonucleotide synthesis. Further, decreasing the number of changes in the nucleotide sequence decreases the potential of altering the secondary structure of the sequence, which can have a significant impact on gene expression in certain host cells. The introduction of undesirable restriction

sites is also reduced, facilitating the subcloning of the genes of interest into the plasmid expression vector.

[0133] In a fourth method, termed "standardized optimization," a Codon Usage Table (CUT) for the sequence to be optimized is generated and compared to the CUT for human genomic DNA (see, e.g., Table 8 below). Codons are identified for which there is a difference of at least 10 percentage points in codon usage between human and query DNA. When such a codon is found, all of the wild type codons for that amino acid are modified to conform to predominant human codon.

[0134] The codon usage frequencies for all established SARS-CoV open reading frames (ORFs) is compared to the codon usage frequencies for humans in Table 8 below.

TABLE 8

SARS CoV Urban Codon Frequencies using all established ORFs

Amino Acid	Codon	Urban Number	Urban Frequency of usage	Human Number	Human Frequency of usage
Phe	UUU	272	0.6154	326146	0.4525
Phe	UUC	170	0.3846	394680	0.5475
Total		442		720826	
Leu	UUA	150	0.1777	139249	0.0728
Leu	UUG	150	0.1777	242151	0.1266
Leu	CUU	254	0.3009	246206	0.1287
Leu	CUC	119	0.1410	374262	0.1956
Leu	CUA	90	0.1065	133980	0.0700
Leu	CUG	81	0.0950	777077	0.4062
Total		844		1912925	
Ile	AUU	262	0.5794	303721	0.3554
Ile	AUC	98	0.2163	414483	0.4850
Ile	AUA	93	0.2053	136399	0.1596
Total		453		854503	
Met	AUG	212	0.0005	430946	1.0000
Total		212		430946	
Val	GUU	299	0.4194	210423	0.1773
Val	GUC	126	0.1767	282465	0.2380
Val	GUA	152	0.2132	134991	0.1137
Val	GUG	136	0.1907	559044	0.4710
Total		713		1186903	
Ser	UCU	202	0.3328	282407	0.1840
Ser	UCC	41	0.0675	336349	0.2191
Ser	UCA	176	0.2900	225963	0.1452
Ser	UCG	20	0.0329	86761	0.0565
Ser	AGU	118	0.1944	230047	0.1409
Ser	AGC	40	0.0824	373362	0.2433
Total		607		1534889	
Pro	CCU	163	0.4405	333705	0.2834
Pro	CCC	38	0.1027	386462	0.3281
Pro	CCA	156	0.4216	322220	0.2726
Pro	CCG	13	0.0351	135317	0.1149
Total		370		1177704	
Thr	ACU	275	0.4264	247913	0.2419
Thr	ACC	86	0.1333	371420	0.3624
Thr	ACA	257	0.3985	285655	0.2787
Thr	ACG	27	0.0419	120022	0.1171
Total		645		1025010	

[0135] The present invention provides isolated polynucleotides comprising codon-optimized coding regions of SARS-CoV polypeptides, e.g., S, S1, S2, N, E, or M, or fragments, variants, or derivatives thereof.

-continued

GCACCGCGCTGCTACCGCCACAGCAGCAGCGGTTCACGCCCTTCACAGAG
 TTCGCGCGGAGAGTGAGGCACTTCACCGACAGCGCTCGGGACCCCAAGAC
 CACGCGAGATCTCGACATCAGCCCTCGAGCTTCGGCGCGCTGAGCGTGA
 CACACCCCGGCACCAACGCGACGAGGAGGTGCGCTGTCTACCGAGGAC
 GTGAACCTGACCGGAGCTGAGCAGCGCCATCCACGCGGACAGCTGACCCC
 CGCGTGGCGAGATCTACGACACCGCAACACGTCTTCAGACCCAGGCGG
 CGTGCCTGATCGCGCGGACGCTGACACAGCTACGAGTGGGACATC
 CCGATCGGCGCGGACCTGCGCGAGCTACACACCTCGAGCGTCTGCTGG
 GAGCAGCAGCCAGAGAGCATGTGGGCTACACCATGAGCCTGGGCGCGG
 ACAGCAGCATCGCTACGACCAACACCATGCCATCCCCCAACTTC
 AGCATCAGCATCACACCGAGGTGATGCCCTGAGCATGGCAGACAGCAG
 CGTGACTGCACATGTATCTCTCGCGGACGACACCGAGTGGCGCAACG
 TGTCTCTGCACTACGCACTTCTGTCACCGAGTGAACCGGCGCTGAGC
 GGCATCGCGCGGACGACGCGAAGACACCGGAGGTGTTCGCCCAAGT
 GAAGCAGATGTACAGGACCCCACTTGAATCTTCGCGGCTTCACT
 TCAGCCAGATCTGCCGACCCCTGAAAGCCACCAAGCGGAGCTTCATC
 GAGGACCTGCTGTCAACAGATGACCTGGCGGAGCGCGCTTCTATGA
 CGAGTAGCGGAGTGTCTGGGAGCATCAACGCCGCGGAGCTGATCTCG
 CCCAGAGTTCAACGGCTGACCTGCTGCCCGCTCTGCTGACGAGCAGC
 ATGATCTGCCGCTACACGCGCGCTGTGTCAGCGGACCGCCACCGCGG
 CTGGACCTCTGGCGCGCGCGCGCTCGAGATCCCTTCGCCATCGAGA
 TGGCTATCGGTTCAACGGCATCGCGTACCCAGAACCTGCTGTACAG
 AACCGAGACGATGTGCCAACGATTCANCAAGGCGCTACGCGAGATCA
 GAGAGCGTACACCAACACGACCGCGCTGGGCAAGCTGCGAGAGCTGG
 TGAACCGAGAACCGGCGCGCTGAACACCTGTGTGAAGCAGCTGACGAGC
 AACTTCTGGCGCATCAGCAGCTGCTGAACGACATCTGACCGCGCTGA
 CAGGTGGAGCGGAGTGTGACATCGACGCGCTGATCAACGGCGGCTG
 AGAGCTTCGACGACTAGCTGACCGACGAGCTGATCTCGGCGCGCGGAGC
 CGGCGCGCGCAACTGCGCGCGCAACAGATGAGCGAGTGGCTGGG
 CGAGAGCAGCGGTTGACTTCTCGCGGCAAGGCTACACCTGATGAGCT
 TCCCGAGCGCGCGCGCGCGCGCTGTGGTGTCTGACGTGACCTACGTG
 CCCAGCGAGGCGGAGACTTCACACCGCGCGCGCATCTGCGCAGAGG
 CAGAGGCTACTTCCCGGAGGCGGTGTGTGTTCACAGGCAACAGCT
 GGTTCATCAACGAGAGACTCTTCACCGCGCGGATCATCAACACGAGC
 AACACCTCTGTGAGCGGCACTGACGAGTGTGATCGGATCATCAACAA
 CACGTGTACGACCCCTGACGCGGAGCTGACAGCTTCAAGGAGGAGC
 TGGACAGTACTTCAAGACCCACACGAGCCGCGAGCTGAGCTGGGCGAC
 ATCAGCGGATCAACCGCGAGCTGGTGAACATCCAGAGGAGATCGAGCG

-continued

GCTGAAGAGTGGCCAGAACTCAAGAGAGCGCTGATGACCTGACG
 AGCTGGGCAAGTAGAGCAGTATCATCAAGGCGCTGG

[0140] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:2 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:2 as follows: about 37 of the 81 phenylalanine codons are TTT, and about 44 of the phenylalanine codons are TTC; about 7 of the 92 leucine codons are TTA, about 12 of the leucine codons are TTG, about 12 of the leucine codons are CTT, about 18 of the leucine codons are CTC, about 7 of the leucine codons are CTA, and about 36 of the leucine codons are CTG; about 26 of the 74 isoleucine codons are ATT, about 35 of the isoleucine codons are ATC, and about 13 of the isoleucine codons are ATA; the 18 methionine codons are ATG; about 15 of the 86 valine codons are GTT, about 40 of the valine codons are GTG, about 10 of the valine codons are GTA, and about 21 of the valine codons are GTC; about 17 of the 91 serine codons are TCT, about 20 of the serine codons are TCC, about 14 of the serine codons are TCA, about 5 of the serine codons are TCG, about 13 of the serine codons are AGT, and about 22 of the serine codons are AGC; about 16 of the 56 proline codons are CCT, about 18 of the proline codons are CCC, about 16 of the proline codons are CCA, and about 6 of the proline codons are CCG; about 23 of the 96 threonine codons are ACT, about 35 of the threonine codons are ACC, about 27 of the threonine codons are ACA, and about 11 of the threonine codons are ACG; about 21 of the 81 alanine codons are GCT, about 33 of the alanine codons are GCC, about 18 of the alanine codons are GCA, and about 9 of the alanine codons are GCG; about 23 of the 52 tyrosine codons are TAT and about 29 of the tyrosine codons are TAC; about 6 of the 14 histidine codons are CAT and about 8 of the histidine codons are CAC; about 14 of the 55 glutamine codons are CAA and about 41 of the glutamine codons are CAG; about 37 of the 81 asparagine codons are AAT and about 44 of the asparagine codons are AAC; about 24 of the 56 lysine codons are AAA and about 32 of the lysine codons are AAG; about 32 of the 70 aspartic acid codons are GAT and about 38 of the aspartic acid codons are GAC; about 17 of the 40 glutamic acid codons are GAA and about 23 of the glutamic acid codons are GAG; about 14 of the 30 cysteine codons are TGT and about 16 of the cysteine codons are TGC; the 10 tryptophan codons are TGG; about 3 of the 39 arginine codons are CGT, about 7 of the arginine codons are CGC, about 4 of the arginine codons are CGA, about 8 of the arginine codons are CGG, about 9 of the arginine codons are AGA, and about 8 of the arginine codons are AGG; and about 12 of the 74 glycine codons are GGT, about 25 of the glycine codons are GGC, about 19 of the glycine codons are GGA, and about 18 of the glycine codons are GGG.

[0141] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must

remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0142] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:2, optimized according to codon usage in humans is presented herein as SEQ ID NO:24.

ATG TTT ATC TTC CTC CTC TTC CTG ACG CTC ACT AGC
GGA TCC GAC TTA GAT CGG TGT ACC ACT TTC GAC GAC
GTC CAG GCC CCT AAC TAT ACT CAA CAT ACC TCC AGT
ATG CGC GGG GTG TAC TAT CCA GAT GAG ATT TTT CGG
AGC GAC ACT CTG TAC TTA ACA CAG GAC CTG TTT CTA
CCG TTT TAT TCA AAT GTA ACC GGC TTC CAC ACC ATT
AAC CAT ACA TTT GGC AAT CCC GTG ATA CCA TTC AAA
GAC GGC ATT TAC TTC GCC GCA ACA GAA AAG AGC AAT
GTT GTG AGG GGG TGG GTC TTC GGC TCC ACA ATG AAC
AAT AAA TCT CAG TCT GTG ATC ATC ATC AAT AAC AGC
ACT AAC GTG GTA ACT CGT GCC TGC AAT TTC GAG CTT
TGT GAC AAC CCA TTC TTC GCC GTG TCT AAG CTT ATG
GAC ACC CAG ACT CAC ACA ATG ATC TTT GAC AAT GCT
TTC AAC TGC ACC TTC GAA TAC ATA TCA GAT GCA TTC
TCT TTG GAT GTC AGT GAA AAG TCT GGA AAC TTT AAA
CAT CTG AGA GAG TTT GTC TTC AAA AAC AAG GAC GGC
TTT CTC TAC GTT TAC AAG GGT TAT CAG CCC ATT GAT
GTG GTG CGG GAC CTC CCT TCA GGG TTT AAC ACA TTG
AAA CCA ATA TTC AAA CTC CCG CTG GGT ATC AAT ATT
ACT AAC TTT CGA GCC ATT TTC ACC GGC TTT TCC CCC
GGC CAA GAC ATA TGG GGA ACC AGC GCG GCA GCC TAT
TTC GTC GGT TAT CTG AAG CCC ACT ACA TTT ATG CTG
AAG TAC GAC GAG AAC GGA ACC ATT ACC GAT GCT GTC
GAT TGT TCA CAG AAT CCA CTG GCT GAA TTG AAA TGC
TCC GTG AAG AGC TTT GAG ATC GAT AAG GGG ATT TAC
CAG ACG TCT AAT TTT CAA GTG GTT CCC TCA GGA GAT
GTG GTT AGA TTC CCC AAT ATC ACA AAT TTG TGC CCC
TTC GGT GAA GTG TTT AAT GCC ACA AAG TTC CCG TCT
GTC TAC GCT TGG GAG CGG AAA AAG ATA AGC AAC TGT
GTC GCG GAT TAC AGT GTC CTA TAT AAC TGC ACC TTT
TTT AGC ACG TTC AAG TGT TAC GGG GTG AGT GCT ACT
AAA CTG AAT GAT TTA TCT TTT AGT AAC GTT TAT GCA
GAC TCC TTT GTT GTA AAG GGT GAT GAC GTG CGC CAA
ATT GCA CTT GGG GAC ACC GGA GTG ATG CCA GAT TAT

-continued

AAC TAC AAA CTT CCA GAC GAC TTT ATG GGA TGC GTG
CTC GCC TGG AAC ACT CGC AAC ATC GAC GCA ACC AGC
ACC GGG AAC TAT AAT TAC AAA TAC AGA TAC CTC AGG
CAC GGC AAG CTG CGG CCT TTT GAG CGG GAT ATC TCA
AAC GTC CCA TTT AGC CGG GAC GGC AAG CCC TGT ACT
CCT CCC GCA CTT AAC TGT TAC TGG CCA CTC AAC GAT
TAT GGC TTT TAT ACC ACA ACC GGC ATC GGC TAC CAG
CCC TAC CGG GTG GTG GTG CTA TCT TTC GAG CTG CTG
AAC GCG COT GCC ACC GTA TGT GGG CCC AAG CTT TCG
ACA GAT CTC ATC AAG AAC CAA TGC GTA AAT TTC AAT
TTC AAT GGC CTT ACA GGA ACC GGT GTG CTG ACA CCC
TCC TCC AAG AGG TTT CAA CTT TTC CAG CAG TTT GGA
CGT GAC GTC TCA GAC TTT ACT GAC AGT GTG AGG GAT
CCT AAG ACC TCT GAA ATC CTG GAT ATA TCT CCG TGT
TCC TTC GGT GGG GTT AGT GTG ATA ACC COT GGG ACA
AAT GCT AGT TGC GAA GTG GCC GTA CTC TAT CAA GAC
GTG AAC TGC ACA GAC GTG TCA ACC GCC ATC CAC GGT
GAT CAA CTC ACA CCG GCT TGG GGC ATC TAT AGC ACT
GGC AAT AAC GTG TTC CAA ACG CAG GCC GGC TCC CTT
ATA GGG GCA GAG CAT GTC GAC ACT TCT TAC GAG TGT
GAT ATA CCA ATC GGA GCC GGC ATC TGC GCC TCA TAC
CAC ACG GTG AGC TTG CTG GGC TCC ACC AGT CAG AAG
AGT ATT GTC GCA TAC ACC ATG TCA CTC GGC GCA GAT
TCA AGT ATC GCC TAC AGC AAT AAC ACT AAT GCT ATT
CCT ACC AAC TTT TCC ATT TCC ATC ACA ACT GAG GTT
ATG COT GTC TCC ATG GCT AAG ACT TCC GTG GAC TGC
AAT ATG TAC ATT TGT GGG GAC TCT ACC GAG TGC GGT
AAC CTT TTA CTG CAG TAT GGC TCC TTC TGC ACA CAG
CTG AAT AGA GCC CTG AGC GGA ATT GCC COT GAG CAG
GAT AGA AAT ACG AGA GAA GTG TTT CCG CAG GTG AAA
CAG ATG TAT AAG ACT CCA ACC TTG AAG TAT TTC GGA
GGG TTC AAT TTT AGC CAG ATC CTT COT GAG CCC TTG
AAG CCG ACC AAA AGG ACC TTC ATC GAA GAT CTT CTG
TTC AAC AAA GTT ACT TTA CCG GAC GCC GGG TTC ATG
AAA CAG TAT GGC GAG TGT CTC GGG GAT AAT AAT GCC
CGC GAT CTC ATC TGT GCT CAG AAA TTC AAC GGC CTC
ACA GTG CTC CCC CCA CTT CTG ACG GAT GAT ATG ATC
GCC GCT TAC ACA GCC GCA CTC GTG AGC GGC ACC GCC

-continued

ACA GCC GGT TGG ACA TTC GGA GCT GGA GCC GCA TTA
 CAG ATT CCA TTC GCT ATG CAG ATG GCG TAC AGG TTC
 AAC GGA ATA GGC GTG ACC CAG AAC GTG TTG TMT GAA
 AAT CAG AAG CAG ATT GCG AAG CAG TTC AAC AAA GCC
 ATT TCT CAA ATC CAG GAG TCC CTG ACC ACC ACA AGC
 ACG GCA CTG GGA AAG GTG CAA GAC GTG GTC AAC CAG
 AAC GCC CAA GCC CTA AAT ACC CTG GTT AAG CAG CTG
 TCT AGC AAT TTT GGA GCG ATT TCA TCT GTC CTT AAC
 GAT ATA CTA TCA AGA CTC GAC AAA GTG GAG GCA GAG
 GTC CAA ATC GAC GGC CTG ATT ACG GGC GGC CTC CAG
 AGC CTT CAG ACG TAT GTG ACA CAG CAG CTG ATA AGA
 GCT GCT GAA ATA GGA GCC TCG GCT AAT CTG GCC GCA
 ACC AAA ATG TCC GAA TGC CTC GTG GGG CAG TCC AAA
 GGT GTC GAT TTC TCG GGC AAA GGT TAC CAT TTG ATG
 TCA TTT CCA CAG GCG GGT CCT CAC GGC GTA GTG TTT
 CTC CAC GTG ACT TAT GTA CTT TCG CAG GAA AGG AAC
 TTC ACA ACT GCC CCA GCC ATC TGC CAT GAG GGA AAA
 GCA TAT TTC CCC CGA GAA GGT GTT TTC GTT TTC AAC
 GGG ACA AGC TGG TTC ATT ACT CAA AGG AAT TTT TTT
 TCG CCA CAG ATC ATT ACC ACT GAT AAC ACA TTT GTA
 TCT GGT AAC TGC CAG GTA GTT ATC GGG ATT ATC AAT
 AAT ACG GTC TAT GAC CCC TTG CAA CCT GAG CTG GAT
 AGC TTT AAG GAA GAG CTG CAG AAG TAC TTT AAG AAT
 CAC ACC TCT CCA GAC GTG GAC CTG GGA GAC ATC TCC
 AGG ATT AAT GCA AGT GTT GTG AAT ATT CAG AAA GAG
 ATT GAT AGA CTA AAC GAA GTT GCT AAG AAC TTG AAT
 GAG AGT TTA ATT CAG CTA CAG GAG CTC GGT AAG TAC
 GAA CAG TAC ATC AAA TGG CCG TGG

[0143] Another representative codon-optimized coding region encoding SEQ ID NO:2 is presented herein as SEQ ID NO: 44.

ATG TTT ATC TTC CTG CTG TTT CTG ACA CTG ACA AGC
 GGC AGT GAC CTG GAT AGA TGC ACA ACG TTT GAC GAC
 GTG CAG GCC CCC AAC TAC ACC CAG CAT ACA TCC AGC
 ATG AGG GGC GTT TAC TAC CCC GAT GAG ATC TTT AGA
 AGT GAT ACT CTG TAT CTG ACT CAG GAC CTG TTT CTG
 CCC TTC TAT TCT AAC GTT ACT GGC TTC CAT ACA ATC
 AAC CAC ACC TTC GGC AAC CCC GTA ATA CCC TTT AAG

-continued

GAT GGC ATC TAC TTT GCC GCC ACC GAG AAG TCT AAC
 GTA GTG AGA GGC TGG GTG TTC GGC AGT ACT ATG AAC
 AAC AAG TCT CAG TCT GTG ATA ATA ATC AAC AAC TCC
 ACT AAC GTC GTC ATC AGA GCC TGT AAC TTC GAG CTG
 TGC GAT AAC CCC TTC TTC GCC GTT TCG AAG CCC ATG
 GGC ACT CAG ACC CAT ACA ATG ATC TTT GAT AAC GCC
 TTC AAC TGC ACC TTT GAG TAT ATC TGC GAT GCC TTC
 AGT CTG GAT GTG TCC GAG AAG TCA GGC AAC TTC AAG
 CAT CTG AGA GAG TTT GTG TTC AAG AAC AAG GAT GGC
 TTT CTG TAC GTC TAC AAG GGC TAC CAG CCC ATA GAT
 GTG GTA CGT GAC CTG CCC AGC GGC TTC AAC ACT CTG
 AAG CCC ATA TTC AAG CTG CCC CTG GGC ATA AAC ATT
 ACC AAC TTT AGA GGC ATT CTG ACG GGC TTC TCC CCC
 GCC CAG GAT ATC TGG GGC ACA AGT GGC GCC GCC TAC
 TTC GTG GGC TAC CTG AAG CCC ACA ACT TTT ATG CTG
 AAG TAC GAC GAG AAC GGC ACC ATA ACA GAT GCC GTG
 GAC TGT TCT CAG AAC CCC CTG GGC GAG CTG AAG TGC
 TCA GTT AAG AGT TTT GAG ATA GAT AAG GGC ATC TAT
 CAG ACA AGC AAC TTC GGC GTG GTC CCC AGC GGC GAT
 GTG CTG AGG TTT CCC AAC ATT ACC AAC CTG TGC CCC
 TTC GGC GAG GTA TTC AAC GCC ACA AAG TTC CCC TCC
 GTT TAC GCC TGG GAG AGG AAG AAT ATT TCA AAC TGC
 GTG GCC GAC TAC TCG GTG CTG TAT AAC TCT ACT TTC
 TTC AGT ACC TTT AAG TGC TAC GGC GTG TCT GGC ACA
 AAG CTG AAC GAT CTG TGC TTT AGC AAC GTG TAT GCC
 CAT AGC TTC GTC GTC AAC GGC GAC GAC GTC AGA CAG
 ATC CCC CCC GGC CAG ACA GGC GTC ATC GGC GAC TAC
 AAC TAC AAG CTG CCC GAC GAT TTC ATG GGC TGC GTG
 CTG GCC TGG AAC ACG ACG AAC ATA GAT GCC ACC AGC
 ACT GGC AAC TAC AAC TAC AAG TAC AGA TAT CTG CGG
 CAC GGC AAG CTG ACG CCC TTC CAG AGA GAC ATC TCT
 AAC GTT CCC TTT TCC CCC GAT GGC AAG CCC TGC ACT
 CCC CCC GCC CTG AAC TGC TAC TGG CCC CTG AAC GAC
 TAT GGC TTC TAC ACC ACA ACT GGC ATC GGC TAT CAG
 CCC TAC CGC GTA CTG GTC CTG TGC TCC AAC CTG CTG
 AAC GCC CCC GGC ACA GTC TGC GGC CCC AAG CTG TCC
 ACT CAG CTG ATT AAG AAC CAG TGT GTG AAC TTC AAC
 TTT AAC GGC CTG ACT GGC ACC GGC GTG CTG ACA CCC
 AGC AGC AAG CGG TTC CAG CCC TTC CAG CAG TTT GGC

-continued

AGA GAC GTG TCT GAT TTC ACA GAT TCC GTG AGA GAT
 CCC AAG ACT TCC GAG ATA CTG GAT ATC AGT CCC TGC
 TCC TTC GGC GGC GTG TCA GTT ATT ACA CCC GGC ACT
 AAC GCC TGG TCC GAG GTA GCC GTT CTG TAT CAG GAC
 GTG AAC TGC ACT GAT GTG AGT ACA GCC ATC CAC GCC
 GAC CAG CTG ACC CCC GGC TGG CGG ATT TAT AGT ACG
 GGC AAC AAC GTC TTT CAG ACT CAG GCC GGC TGC CTG
 ATC GGC GCC GAG CAT GTA GAT ACG TCT TAT GAG TGC
 GAC ATC CCC ATC GGC GCC GGC ATC TGC GCC AGC TAT
 CAC ACC GTT TCT CTG CTG CGA AGT ACT TCT CAG AAG
 TCT ATA GTG GCC TAC ACC ATG TCT CTG GGC GCC GAT
 AGC TCT ATC GCC TAT AGC AAC AAC ACT ATA GCC ATC
 CCC ACA AAC TTC TCT ATT TCT ACT ACT ACA GAG GTG
 ATG CCC GTC TCC ATG GCC AAG ACC AGC GTT GAT TGC
 AAC ATG TAC ATC TGC GGC GAT AGT ACA GAG TGC GGC
 AAC CTG CTG CTG CAG TAT GGC AGC TTC TGC ACC CAG
 CTG AAC AGA GCC CTG TCT GGC ATC GCC GCC GAG CAG
 GAT AGG AAC ACA AGA GAG GTT TTC GCC CAG GTT AAG
 CAG ATG TAC AAG ACT CCC ACT CTG AAG TAC TTT GGC
 GGC TTT AAC TTT TCT CAG ATT CTG CCC GAT CCC CTG
 AAG CCC ACT AAG AGG AGT TTC ATA GAG GAC CTG CTG
 TTC AAC AAG GTG ACT CTG GCC GAC GCC GGC TTT ATG
 AAC CAG TAC GGC CAG TGC CTG GGC GAT ATC AAC GCC
 AGA GAC CTG ATC TGT GGC CAG AAG TTT AAC GGC CTG
 ACA GTA CTG CCC CCC CTG CTG ACT GAT GAC ATG ATT
 GCC GCC TAT ACG GCC GGC CTG GTG TCT GGC ACT GGC
 ACC GCC GGC TGG ACC TTT GGC GCC GGC GCC GGC CTG
 CAG ATA CCC TTT GGC ATG CAG ATG GGC TAC CGA TTC
 AAC GGC ATA GGC GTA ACC CAG AAC GTT CTG TAT GAG
 AAC CAG AAG CAG ATA GCC AAC CAG TTC AAC AAG GCC
 ATC TCT CAG ATT CAG GAG TCT CTG ACC ACT ACA TCT
 ACT GCC CTG GGC AAG CTG CAG GAG GTA GTG AAC CAG
 AAC GCC CAG GCC CTG AAC ACC CTG GTT AAG CAG CTG
 TCA AGT AAC TTC GGC GCC ATC TCT AGC GTT CTG AAC
 GAT ATA CTG AGT CGC CTG GAT AAG GTG GAG GCC GAG
 CTG CAG ATT GAC AGA CTG ATC ACA GGC AGA CTG CAG
 TCT CTG CAG ACA TAT GTT ACT CAG CAG CTG ATA AGG
 GCC GCC GAG ATT AGA GCC AGT GCC AAC CTG GCC GCC

-continued

ACT AAG ATG TCC GAG TGC GTC CTG GGC CAG AGT AAG
 AGG GTA GAC TTT TGT GGC AAG GGC TAT CAC CTG ATG
 TCC TTC CCC CAG GCC GGC CCC CAC GGC GTC GTG TTT
 CTG CAT GTC ACT TAT GTT CCC TCA CAG GAG AGG AAC
 TTC ACG ACC GCC CCC GGC ATC TGC CAG CAG GGC AAG
 GCC TAT TTC CCC AAG GAG GGC GTC TTC GTA TTC AAC
 GGC ACG AGT TGG TTC ATC ACC CAG CGA AAC TTC TTT
 TCG CCC CAG ATA ATT ACA ACG GAC AAC ACT TTT GTA
 AGT GGC AAC TGC GAT GTC GTC ATC GGC ATA ATC AAC
 AAC ACG GTT TAC GAC CCC CTG CAG CCC GAG CTG GAT
 TCA TTC AAG GAG GAG CTG GAC AAG TAC TTC AAG AAC
 CAT ACT AAG CCC GAC GTT GAT CTG GGC GAC ATA AGC
 GGC ATC AAC GCC AGT GTA GTC AAC ATA CAG AAG GAG
 ATC GAT AGA CTG AAC GAG GTG GCC AAC AAG CTG AAC
 GAG TCT CTG ATA GAC CTG CAG GAG CTG GGC AAG TAC
 GAG CAG TAC ATC AAG TGG CCC TGG

[0144] A representative codon-optimized coding region encoding SEQ ID NO:2 according to the "standardized optimization" method is presented herein as SEQ ID NO: 67.

ATG TTC ATC TTC CTG CTG TTC CTG ACC CTG ACC AGC
 GGC AGC GAC CTG GAT CGC TGC ACC ACC TTC GAT GAC
 GTG CAG GCC CCC AAC TAC ACC CAG CAT ACC AGC AGC
 ATG CGC GGC GTG TAC TAC CCC GAT GAG ATC TTC CGC
 AGC GAC ACC CTG TAC CTG ACC CAG GAC CTG TTC CTG
 CCC TTC TAC AGC AAC GTG ACC GGC TTC CAC ACC ATC
 AAC CAT ACC TTC GGC AAC CCC CTG ATC CCC TTC AAG
 GAC GGC ATC TAC TTC GGC GCC ACC GAG AAG AGC AAC
 GTG GTG CGC GGC TGG GTG TTC GGC ACC ACC ATG AAC
 AAC AAG AGC CAG AGC GTG ATC ATC ATC AAC AAC AGC
 ACC AAC GTG GTG ATC CGC GCC TGC AAC TTC GAG CTG
 TGC GAC AAC CCC TTC TTC GCC GTG AAG AAC CCC ATG
 GGC ACC CAG ACC CAT ACC ATG ATC TTC GAT AAC GCC
 TTC AAC TGC ACC TTC GAG TAC ATC AGC GAC GCC TTC
 AGC CTG CAG GTG AGC GAG AAG AGC GGC AAC TTC AAG
 CAT CTG CGC GAG TTC GTG TTC AAG AAC AAG GAT GGC
 TTC CTG TAC GTG TAC AAG GGC TAC CAG CCC ATC GAC
 GTG GTG CGC GAT CTG CCC AGC GGC TTC AAC ACC CTG
 AAG CCC ATC TTC AAG CTG CCC CTG GGC ATC AAC ATC

-continued

ACC AAC TTC CGC GCC ATC CTG ACC GCC TTC AGC CCC
 GCC CAG GAC ATC TGG GGC ACC AGC GCC GCC TAC
 TTC GTG GGC TAC CTG AAC CCC ACC ACC TTC ATG CTG
 AAG TAC GAT GAG AAC GGC ACC ATC ACC GAC GCC GTG
 GAC TGC AGC CAG AAC CCC CTG GCC GAG CTG AAG TGC
 AGC GTG AAG AGC TTC GAG ATC GAT AAG GGC ATC TAC
 CAG ACC AGC AAC TTC CGC GTG CTG CCC AGC GGC GAC
 GTG GTG CGC TTC CCC AAC ATC ACC AAC CTG TGT CCC
 TTC GGC GAG GTG TTC AAC GCC ACC AAG TTC CCC AGC
 GTG TAC GCC TGG GAG CGC AAG AAG ATC AGC AAC TGC
 GTG GCC GAC TAC AGC GTG CTG TAC AAC AGC ACC TTC
 TTC AGC ACC TTC AAG TGC TAC GGC GTG AGC GCC ACC
 AAG CTG AAC GAT CTG TGC TTC AGC AAC GTG TAC GCC
 GAC AGC TTC GTG GTG AAG GGC GAT GAT GTG CGC CAG
 ATC GCC CCC GGC CAG ACC GGC GTG ATC GCC GAT TAC
 AAC TAC AAG CTG CCC GAC GAC TTC ATG GGC TGC GTG
 CTG GCC TGG AAC ACC CGC AAC ATC GAC GCC ACC AGC
 ACC GGC AAC TAC AAC TAC AAG TAC CGC TAC CTG CGC
 GAT GGC AAG CTG CGC CCC TTC GAG CGC GAT ATC AGC
 AAC GTG CCC TTC AGC CCC GAT GGC AAG CCC TGC ACC
 CCC CCC GCC CTG AAC TGT TAC TGG CCC CTG AAC GAC
 TAC GGC TTC TAC ACC ACC ACC GGC ATC GGC TAC CAG
 CCC TAC CGC GTG GTG GTG CTG AGC TTC GAG CTG CTG
 AAC GCC CCC GCC ACC GTG TGC GGC CCC AAG CTG AGC
 ACC GAC CTG ATC AAG AAC CAG TGC GTG AAC TTC AAC
 TTC AAC GGC CTG ACC GGC ACC GGC GTG CTG ACC CCC
 AGC AGC AAG CGC TTC CAG CCC TTC CAG CAG TTC GGC
 GGC GAT GTG AGC GAC TTC ACC GAT AGC GTG CGC GAC
 CCC AAG ACC AGC GAG ATC CTG GAT ATC AGC CCC TGC
 AGC TTC GGC GGC GTG AGC GTG ATC ACC CCC GGC ACC
 AAC GCC AGC AGC GAG GTG GCC GTG CTG TAC CAG GAT
 GTG AAC TGT ACC GAT GTG AGC ACC GCC ATC CAC GCC
 GAT CAG CTG ACC CCC GGC TGG CGC ATC TAC AGC ACC
 GGC AAC AAC GTG TTC CAG ACC CAG GCC GGC TGC CTG
 ATC GGC GCC GAG CAT GTG GAC ACC AGC TAC GAG TGT
 GAC ATC CCC ATC GGC GCC GGC ATC TGT GCC AGC TAC
 CAC ACC GTG AGC CTG CTG CGC AGC ACC AGC CAG AAG
 AGC ATC GTG GCC TAC ACC ATG AGC CTG GGC GCC GAT

-continued

AGC AGC ATC GCC TAC AGC AAC AAG ACC ATC GCC ATC
 CCC ACC AAC TTC AGC ATC AGC ATC ACC ACC GAG GTG
 ATG CCC GTG AGC ATG GCC AAG ACC AGC GTG GAC TGC
 AAC ATG TAC ATC TGC GGC GAT AGC ACC GAG TGC CCC
 AAC CTG CTG CTG CAG TAC GGC AGC ACC GGC ACC CAG
 CTG AAC CGC GGC CTG AGC GGC ATC GCC GGC GAG CAG
 GAT CGC AAC ACC CGC GAG GTG TTC GCC CAG GTG AAG
 CAG ATG TAC AAG ACC CCC ACC CTG AAG TAC TTC GGC
 GGC TTC AAC TTC AGC CAG ATC CTG CCC GAT CCC CTG
 AAG CCC ACC AAG CGC AGC TTC ATC GAG GAT CTG CTG
 TTC AAC AAG GTG ACC CTG GCC GAT GCC GGC TTC ATG
 AAG CAG TAC GGC GAG TGC CTG GGC GAT ATC AAC GGC
 CGC GAT CTG ATC TGC GCC CAG AAG TAT AAC GGC CTG
 ACC GTG CTG CCC CCC CTG CTG ACC GAC GAC ATG ATC
 GCC GCC TAC ACC GCC GCC CTG GTG AGC GGC ACC GCC
 ACC GCC GGC TGG ACC TTC GGC GCC GGC GCC GCC CTG
 CAG ATC CCC TTC GCC ATG CAG ATG GCC TAC CGC TTC
 AAC GGC ATC GGC CTG ACC CAG AAC GTG CTG TAC GAG
 AAC CAG AAG CAG ATC GCC AAC CAG TTC AAC AAG GCC
 ATC AGC CAG ATC CAG GAG AGC CTG ACC ACC ACC AGC
 ACC GCC CTG GGC AAG CTG CAG AAG GTG GTG AAC CAG
 AAC GCC CAG GCC CTG AAC ACC CTG GTG AAG CAG CTG
 AGC AGC AAC TTC GGC GCC ATC AGC AGC GTG CTG AAC
 GAC ATC CTG AGC CGC CTG GAT AAG GTG GAG GCC GAG
 GTG CAG ATC GAT CGC CTG ATC ACC GGC CGC CTG CAG
 AGC CTG CAG ACC TAC GTG ACC CAG CAG CTG ATC CGC
 GCC GCC GAG ATC CGC GCC AGC GCC AAC CTG GCC GGC
 ACC AAG ATG AGC GAG TGC GTG CTG GGC CAG AGC AAG
 CGC GTG GAT TTC TGC GGC AAG GGC TAC CAC CTG ATG
 AGC TTC CCC CAG GCC GCC CCC CAT GGC GTG GTG TTC
 CTG CAC GTG ACC TAC GTG CCC AGC CAG GAG GGC AAC
 TTC ACC ACC GCC CCC GCC ATC TGC CAC GAG GGC AAG
 GCC TAC TTC CCC CGC GAG GGC GTG TTC GTG TTC AAC
 GGC ACC AGC TGG TTC ATC ACC CAG CGC AAC TTC TTC
 AGC CCC CAG ATC ATC ACC ACC GAT AAC ACC TTC GTG
 AGC GGC AAC TGC GAT GTG GTG ATC GGC ATC ATC AAC
 AAC ACC GTG TAC GAT CCC CTG CAG CCC GAG CTG GAC
 AGC TTC AAG GAG GAG CTG GAT AAG TAC TTC AAG AAC
 CAC ACC AGC CCC GAC GTG GAT CTG GGC GAT ATC AGC

-continued

GCG ATC AAC GCC AGC GTG GTG AAC ATC CAG AAG GAG
 ATC GAT CGC CTG AAC GAG GTG GCC AAG AAC CTG AAC
 GAG AGC CTG ATC GAC CTG CAG GAG CTG GGC AAG TAC
 GAG CAG TAC ATC AAG TGG CCG TGG

[0145] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:4 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:4 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:4, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:4 is shown in Table 10.

TABLE 10

AMINO ACID		Number in SEQ ID NO: 4
A	Ala	38
R	Arg	23
C	Cys	20
G	Gly	44
H	His	9
I	Ile	38
L	Leu	46
K	Lys	31
M	Met	5
F	Phe	53
P	Pro	37
S	Ser	56
T	Thr	58
W	Trp	6
Y	Tyr	35
V	Val	53
N	Asn	46
D	Asp	44
Q	Gln	21
S	Glu	17

[0146] Using the amino acid composition shown in Table 10, a human codon-optimized coding region which encodes SEQ ID NO:4 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:4 as follows: the 53 phenylalanine codons are TTC, the 46 leucine codons are CTG, the 38 isoleucine codons are ATC, the 8 methionine codons are ATG, the 53 valine codons are GTG, the 56 serine codons are AGC, the 37 proline codons are CCC, the 58 threonine codons are ACC, the 38 alanine codons are GCC, the 35 tyrosine codons are TAC, the 9 histidine codons are CAC, the 21 glutamine codons are CAG, the 46 asparagine codons are AAC, the 31 lysine codons are AAG, the 44 aspartic acid codons are GAC, the 17 glutamic acid codons are GAG, the 20 cysteine codons are TGC, the 6 tryptophan codons are TGG, the 23 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 44 glycine codons are GGC.

The codon-optimized S1 coding region designed by this method is presented herein as SEQ ID NO:27.

ATGTTTCATCTTCTGCTGTCTTCTGACGCTGACGCGGACGGAGCTGGA
 CAGATGACCCACTTTCGAGAGGCTGACGGCCCACTACACCGAGCGCA
 CCAGCAGCATGAGAGGCTGTACTACCCGACGAGATCTTCAGAACGAC
 ACCCTGTACTGACCCAGGAGCTGTCTCTGCTCTTACAGCAACGTGAC
 CGGCTTCCACACATCAACACACACTTCCGCAACCCCGTATCCCTTCA
 AGGAGGCGATCTACTTCCGCCACCGAGAGGACCACTGTGTGAGAGGC
 TGGTGTTCGGCAGCAGCATGAAACAAAGGCCAGAGCTGATCATAT
 CACACAGCAGCCAGCTGTGTGATCAGAGCTCACTTCTGAGCTGTGGC
 ACACCCCTTCTTCCGCTGAGCAGCCCATGGGACCCAGCCACACAC
 ATGATCTTCGACAAAGCTTCACTGACCTTGGAGTACATCAGCAGCG
 CTCAGCTGTGAGCTGAGCGAGAGAGCGGCALCTTCAAGCAGCTGAGAG
 AGTCTGTGTTCAAGAACAGGACGGCTCTGTCTGCTGACAGAGGCTAC
 CAGCCCATGAGCTGTGTGAGAGAGCTGCGCGCGCTTCAACACCTGAA
 GCGCATCTTCAAGCTGCGCTCGGCTATCACTACCAACTCTAGAGGCA
 TCTTGACCGCTCTCAGCCCGCCGAGGATCTGGGCGACAGCGCGGCC
 CGCTACTTCTGTGGCTACTTGAAGCCGACCACTTCACTGTGAGTACGA
 CGAGAACCGCACCATCAGCAGCGCTGTGATCGACCGAACCCCTTGG
 CCGAGCTGAAGTGCAGCGTGAAGAGCTTCGAGATCGACAGGGGATCTAC
 CAGACCGACCACTTCAAGTGTGTCCCGCGCGCTGTGTGAGTATCCC
 CACATCACCACCACTGTGCGCTTGTGCGGAGTGTTCACGCCACCAAGT
 TCCCGACGCTGATCGCTGGGAGAGAGAGATCAACACTTGTGTGGCC
 GACTACAGCGTGTGTGACACAGCAGCTTCTTCAAGCATCTTCAAGTCTA
 CGCGTGTGAGCGCCACAGCTGAAGCAGCTGTGTCTTACAGCACTGTACG
 CCGACAGCTTCTGTGTGAAGGCGGAGCGTGTGAGCAGATTCGCCCGCGC
 CAGACCGCGCTGATCGCGACTCAACTACAGCTGCCGAGGACTTCACT
 GGGCTGTGCTGTGCGCTGAGAACAGAGAACATCAAGCCACAGCAGCG
 GCACTACCACTCAAGTACAGATACCTGAGACAGCGGCAAGCTGAGACCC
 TTGGAGAGAGATCAGCAAGCTGCCCTTACGCCCGGAGCGGACCGCTG
 CACCCCCCGCCCTGAAGTCTACTGGCCCTCAAGGACTACGGCTTCT
 ACACCAACCGCGCTGCTACCGCCCTACAGCGCTGAGTGTGTGTGAGC
 TTGAGCTGTGTGACGCCCGCCAGCTGTGTGCGGCCGCAAGTGAAGAC
 CGACCTGATCAAGAACCGCTGTGTGACTTCACTTCAACGGCTGACCG
 GCACCGGCTGTGACCCCGCAGCAGCAAGAGATTCGACCGCTTCCAGAG
 TTGCGCAGAGAGCTGAGCGACTTCACGACAGGCTGAGAGACCCAGAC
 CAGCGAGATCTGTGACATCGCCCTGCACTTCTGCGGCGTGTGAGGCTGA
 TCAACCCCGCAGCAACCGCAGCAGGAGGTTGGCTGTGTATCCAGAGAC
 GTGAGCTGCAGCGAGTGAAGCAGCCGCTACGACCGCGACGCTGACCCC

—continued—

CGCGTGGAGATCTACAGCAGCGGCACACGCTTCCAGAGCCAGGCCG
 GCGTGGTATCGGCGCGAGCACTGAGACAGCTACGAGTGCAGATC
 ACCATCGGCGCGGCACTGCGCCAGCTCCACACCCCTGAGCGTCTGAG
 CAGCAGCGCGGAGAGAGCTGCGGCTTACACCATGAGCTCGGCGCGC

[0147] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:4 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:4 as follows: about 24 of the 53 phenylalanine codons are TTT, and about 29 of the phenylalanine codons are TTC; about 3 of the 46 leucine codons are TTA, about 6 of the leucine codons are TTG, about 6 of the leucine codons are CTT, about 9 of the leucine codons are CTC, about 4 of the leucine codons are CTA, and about 18 of the leucine codons are CTG; about 13 of the 38 isoleucine codons are ATT, about 18 of the isoleucine codons are ATC, and about 7 of the isoleucine codons are ATA; the 8 methionine codons are ATG; about 10 of the 53 valine codons are GTT, about 13 of the valine codons are GTC, about 5 of the valine codons are GTA, and about 25 of the valine codons are GTG; about 10 of the 56 serine codons are TCT, about 12 of the serine codons are TCC, about 8 of the serine codons are TCA, about 3 of the serine codons are TCG, about 9 of the serine codons are AGT, and about 14 of the serine codons are AGC; about 10 of the 37 proline codons are CCT, about 12 of the proline codons are CCC, about 11 of the proline codons are CCG, and about 4 of the proline codons are ACT; about 14 of the 58 threonine codons are ACT, about 21 of the threonine codons are ACC, about 16 of the threonine codons are ACA, and about 7 of the threonine codons are ACG; about 10 of the 38 alanine codons are GCT, about 15 of the alanine codons are GGC, about 9 of the alanine codons are GCA, and about 4 of the alanine codons are GCG; about 15 of the 35 tyrosine codons are TAT and about 20 of the tyrosine codons are TAC; about 4 of the 9 histidine codons are CAT and about 5 of the histidine codons are CAC; about 5 of the 21 glutamine codons are CAA and about 16 of the glutamine codons are CAG; about 21 of the 46 asparagine codons are AAT and about 25 of the asparagine codons are AAC; about 13 of the 31 lysine codons are AAA and about 18 of the lysine codons are AAG; about 20 of the 44 aspartic acid codons are GAT and about 24 of the aspartic acid codons are GAC; about 7 of the 17 glutamic acid codons are GAA and about 10 of the glutamic acid codons are GAG; about 9 of the 20 cysteine codons are TGT and about 11 of the cysteine codons are TGC; the 6 tryptophan codons are TGG; about 2 of the 23 arginine codons are CGT, about 4 of the arginine codons are CGC, about 3 of the arginine codons are CGA, about 5 of the arginine codons are CGG, about 4 of the arginine codons are AGA, and about 5 of the arginine codons are AGG; and about 7 of the 44 glycine codons are GGT, about 15 of the glycine codons are GGC, about 11 of the glycine codons are GGA, and about 11 of the glycine codons are GGG.

[0148] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be

understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a given amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0149] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:4, optimized according to codon usage in humans is presented herein as SEQ ID NO:26.

ATG TTT ATC TTT TTG CTG TTT CTC ACA TTA ACT TCG
 GGG TCT GAC CTG GAC CGG TGC ACC ACA TTC GAT GAC
 GTG CAA GCC CCC AAC TAC ACT GAG CAT ACA TCT AGC
 ATG CGC GGG GTG TAC TAC CCA GAT GAG ATC TTT AGG
 TCC GAC ACC CTT TAT CTG ACC GAC GAG CTT TTT CTT
 CCT TTC TAC TCT AAT GTA ACT GGG TTC CAT ACC ATC
 AAC CAT ACC TTT GGC AAC CCA GTG ATT CCA TTT AAG
 GAT GGT ATT TAC TTC GCC CGC ACC GAG AAA TCA AAT
 GTT GTG CGC GGC TGG GTT TTC GGC TCC ACC ATG AAC
 AAT AAG AGT CAG TCC GTA ATT ATT ATT AAC AAT AGT
 ACA AAG GTG GTG ATC AGG GCA TGT AAT TTT GAA TGG
 TGC GAC AAC COT TTT TGC GCT GTA AAG AAA CCC ATG
 GGG ACG GAC ACT CAC AGG ATG ATC TTC GAT AAC GCT
 TTC AAT TGC ACG TTT GAG TAC ATA TCC GAT GGC TTT
 TCT GTA GAT GTG TCC GAA AAA TCA GGG AAT TTT AAG
 CAC CTG AGA GAG TCT GTC TTT AAG AAC AAG GAG GGT
 TTC TTG TAC GTG TAC AAG GGA TAC CAG CCG ATC GAC
 GTG GTG CCG GAC CTA CCC AGC GGA TTC AAC ACC CTC
 AAG CCC ATT TTT AAG CTC CCA GTG GGT ATC AAT ATA
 ACT AAC TTC AGA GCT ATT CTC CAC GAG GGT TTC TCT CAA
 GCT CAG GAT ATT TGG GGG ACT AGT GCG GCA GCT TAT
 TTC GTG GGA TAC CTT AAG CCC ACA ACC TTC ATG TTG
 AAA TAC GAT GAG AAC GGA ACC ATA ACT GAC GCA GTT
 GAC TGC TCA CAG AAC CCC CTC CCA GAG TTG AAA TGC
 TCA GTT AAA TCC TTT GAG GCT GTC AAG GGT ATT TAC
 CAG ACC AGT AAT TTC AGA GTC GTG CCG TCA GGC GAC
 GTC GTG AAG TTT CTT AAC ATC ACA AAT CTA TGT CCT
 TTC GGA GAA GTG TTC ART GCC ACA AAG TTC CCC AGC
 GTG TAC GCC TGG GAG GCA AAA AAG ATA TCT AAT TGC
 GTC GCA GAC TAC AGC GTA CTG TAT AAC AGC ACT TTT
 TTC AGC ACC TTT AAG TGT TAT GGG GTG TCA GCA ACA
 AAA CTG AAC GAT CTC TGC TTT TCA AAC GTT TAT GCC
 GAT TCC TTT GTT GTC AAG GGA GAG GAT GTC COT CAA

-continued

ATT GCT CCC GGG CAA ACT GGC GTT ATC GCT GAC TAT
 AAC TAT AAA CTG CCA GAC GAT TTT ATG GGG TOT GTC
 CTC GCA TGG AAT ACG GGC AAC ATC GAT GCG ACC TCT
 ACC GGA AAC TAC AAC TAT AAA TAT AGG TAT CTT CGG
 CAC GGG AAA TTA CGG CCC TTC GAG CGA GAT ATT TCG
 AAC GTG CCT TTC AGT CCC GAT GGA AAA CCA TOT ACT
 CCT CCA GCC CTC AAT TOT TAT TGC CCA TTG AAT GAC
 TAC GGG TTC TAC ACG ACA ACT GGA ATA GGC TAT CAG
 CCT TAT CGT GTC CTC GTT CTT TCT TTC GAA CTG CTG
 AAT GCT CCC GCC ACG GTG TGC GGT CCA AAA CTC AGC
 ACC GAC CTG ATC AAG AAT CAG TGC GTG AAT TTC AAT
 TTC AAC GGC CTG ACA GGC ACA GGC GTT CTG ACC CCA
 AGC TCC AAG CGC TTC CAG CCC TTC CAG CAA TTT GGC
 AGG GAT GTG TCC GAC TTT ACC GAT TCA GTG CGA GAT
 CCC AAG ACC AGT GAA ATA CTA GAC ATT TCT CGG TGT
 AGC TTT GGC GGC GTG TCT GCT ATT ACT CCT GGG AGG
 AAT GCC TCG AGC GAG GTG GCG GTG TTA TAT CAG GAC
 GTT AAT TGT ACA GAC GTC AGT ACC GCC ATA CAT GCT
 CAT CAG CTG ACT CCT GCA TGG AGA ATC TAC TCC ACA
 GGA AAT AAT GTG TTT CAG ACA CAA GCA GGT TGC CTG
 ATC GGA GCC GAA CAC GTC GAC ACC AGC TAC GAA TGT
 GAT ATC CCT ATC GGT GCC GGC ATC TGC GCT AGT TAT
 CAC ACA GTA AGC CTG CTG CGG AGC ACC AGT CAG AAG
 TCC ATT GTG GCC TAT ACT ATG TCC CTG GGC GCC

[0150] Another representative codon-optimized coding region encoding SEQ ID NO:4 is presented herein as SEQ ID NO:45.

ATG TTC ATC TTC CTG CTG TTT CTG ACA CTG ACT TCT
 GGC TCA GAT CTG GAT AGA TGC ACT ACC TTT GAC GAT
 GTA CAG GCC CCC AAC TAC ACT CAG CAC ACA TCG TCC
 ATG CGA GGC GTG TAT TAC CCC GAC GAG ATC TTC AGA
 AGT GAC ACT CTG TAC CTG ACA CAG GAC CTG TTC CTG
 CCC TTT TAC TCT AAC GTG ACT GGC TTT CAC ACT ATC
 AAC CAT ACC TTC GGC AAC CCC GTA ATC CCC TTC AAG
 GAT GGC ATC TAT TTT GCC GCC ACC GAG AAG TCC AAC
 GTG GTG AGG GGC TGG GTC TTC GGC AGT ACG ATG AAC
 AAC AAG TCT CAG TCC GTG ATA ATC ATA AAC AAC AGT
 ACT AAC GTG GTT ATA AGA GCC TGC AAC TTC CAG CTG

-continued

TGC GAC AAC CCC TTC TTC GGC GTG TCC AAG CCC ATG
 GGC ACA CAG ACC CAC ACC ATG ATA TTC GAC AAC GCC
 TTT AAC TGT ACT TTC GAG TAT ATA AGC GAT GGC TTC
 AGT CTG GAT GTT TCT GAG AAG TCA GGC AAC TTT AAG
 CAT CTG AGA GAG TTC GTA TTC AAG AAC AAG GAC GGC
 TTT CTG TAT GTT TAT AAG GGC TAC CAG CCC ATA GAT
 GTC CTG CGG GAT CTG CCC AGC GGC TTC AAC ACA CTG
 AAG CCC ATT TTT AAG CTG CCC CTC GGC ATC AAC ATA
 ACC AAC TTT AGA GCC ATC CTG ACT GCC TTT AGC CCC
 GCC CAG GAT ATA TGG GGC ACT AGC GCC GCC GCC TAT
 TTC GTC GGC TAC CTG AAG CCC ACC ACA TTC ATG CTG
 AAG TAC GAT AGA AAC GGC ACA ATT ACG GAT GGC GTA
 GAT TGC AGT CAG AAC CCC CTG GCC GAG CTG AAG TGC
 AGT GTG AAG TCT TTC GAG ATC CAG AAG GGC ATA TAC
 CAG ACT TCT AAC TTT CGG GTG GTT CCC AGC GGC GAC
 GTT GTT AGG TTT CCC AAC ATC ACC AAC CTG TGC CCC
 TTC GGC GAG GTG TTT AAC GCC ACA AAG TTC CCC TCC
 GTA TAT GCC TGG GAG AGG AAG AAT TCG AAC TGC
 GTG GCC GAC TAT AGC GTC CTG TAC AAC TCT ACA TTC
 TTT TCT ACA TTC AAG TGC TAC GGC GTC AGT GCC ACT
 AAG CTG AAC GAC CTG TGC TTC AGC AAC GTG TAT GCC
 GAC TCA TTT GTA GTT AAG GGC GAT GAT GTG AGA CAG
 ATT GCC CCC GGC CAG ACA GGC GTG ATC GCC GAT TAT
 AAC TAT AAG CTG CCC GAC GAT TTC ATG GGC TGC GTT
 CTG GCC TGG AAC ACA AGG AAC ATC GAT GCC ACT AGC
 ACT GGC AAC TAC AAC TAC AAG TAC AGG TAT CTG AGA
 CAC GGC AAG CTG AGG CCC TTC GAG CGA GAT ATC AGT
 AAC GTA CCC TTC AGT CCC GAC GGC AAG CCC TGC ACT
 CCC CCC GCC CTG AAC TGC TAT TGG CCC CTG AAC GAC
 TAC GGC TTT TAT ACC ACT ACA GGC ATC GGC TAC CAG
 CCC TAC AGG GTT GTG GTG CTG AGC TTT GAC CTG CTG
 AAC GCC CCC GCC ACT GTT TGC GGC CCC AAG CTG TCA
 ACG GAT CTG ATC AAG AAC CAG TGC GTA AAC TTT AAC
 TTT AAC GGC CTG ACA GGC ACA GGC GTC CTG ACT CCC
 TCT AGT AAG AGA TTC CAG CCC TTT CAG CAG TTC GGC
 CGC GAC GTC AGC GAT TTT ACG GAT AGT GTG AGA GAT
 CCC AAG ACC AGC GAG ATC CTG GAC ATT AGT CCC TGT
 TCT TTC GGC GGC GTG TCT GTC ATA AGC CCC GGC ACG

-continued

AAC GCC TCT TCT GAG GTC GCC GPT CTG TAC CAG GAC
 GTC AAC TGT ACA GAG TTC GCC ACA GCC ATA CAC GCC
 GAT CAG CTG ACT CCC GCG TGG AGA ATT TAC TCT ACC
 GGC AAC AAC GTC TTC CAG ACC CAG GCC GGC TGC CTG
 ATC GGC GCC GAG CAT GTG GAT ACT TCC TAC GAG TGC
 GAC ATA CCC ATC GGC GCC GGC ATT TGC GCC TCG TAC
 CAT ACC GTG TCT CTG CTG AGA TCT ACC TCT CAG AAG
 AGT ATC GTT GCC TAC ACT ATG TCC CTG GGC GCC

[0151] A representative codon-optimized coding region encoding SEQ ID NO:4 according to the "standardized optimization" method is presented herein as SEQ ID NO: 68.

ATG TTC ATC TTC CTG CTG TTC CTG ACC CTG ACC AGC
 GGC AGC GAT CTG GAC CCG TCC ACC ACC TTC GAC GAT
 GTG CAG GCC CCC AAC TAC ACC CAG CAC ACC AGC AGC
 ATG CGC GGC GTG TAC TAC CCC GAT GAG ATC TTC CGC
 AGC GAT ACC CTG TAC CTG ACC CAG GAT CTG TTC CTG
 CCC TTC TAC AGC AAC GTG ACC GGC TTC CAT ACC ATC
 AAC CAC ACC TTC GGC AAC CCC GTG ATC CCC TTC AAG
 GAT GGC ATC TAC TTC GCC GGC ACC GAG AAG AGC AAC
 GTG GTG CGC GGC TGG GTG TTC GGC AGC ACC ATG AAG
 AAC AAG AGC CAG AGC GTG ATC ATC ATC AAC AAC AGC
 ACC AAC GTG GTG ATC CGC GCC TGC AAC TTC GAG CTG
 TGC GAC AAC CCC TTC TTC CCG GTG AGC AAG CCC ATG
 GGC ACC CAG ACC CAC ACC ATG ATC TTC GAC AAC GCC
 TTC AAC TGC ACC TTC GAG TAC ATC AGC GAT GCC TTC
 AGC CTG GAC GTG AGC GAG AAG AGC GGC AAC TTC AAG
 CAT CTG GGC GAG TTC GTG TTC AAG AAC AAG GAT GGC
 TTC CTG TAC GTG TAC AAG GGC TAC CAG CCC ATC GAC
 GTG GTG GGC GAC CTG CCC AGC GGC TTC AAC ACC CTG
 AAG CCC ATC TTC AAG CTG CCC CTG GGC ATC AAC ATC
 ACC AAC TTC CGC GGC ATC CTG ACC GGC TTC AGC CCC
 GCC CAG GAT ATC TGG GCC ACC AGC GCC GCC GGC TAC
 TTC GTG GGC TAC CTG AAG CCC ACC ACC TTC ATG CTG
 AAG TAC GAC GAG AAC GGC ACC ATC ACC GAT GGC GTG
 GAT TGC AGC CAG AAC CCC CTG GGC GAG CTG AAG TGC
 AGC GTG AAG AGC TTC GAG ATC GAT AAG GGC ATC TAC
 CAG ACC AGC AAC TTC CGC GTG GTG CCC AGC GGC GAC
 GTG GTG CGC TTC CCC AAC ATC ACC AAC CTG TGC CCC

-continued

TTC GGC GAG GTG TTC AAC GCC ACC AAG TTC CCC AGC
 GTG TAC GCC TGG GAG CGC AAG AAG ATC AGC AAC TGC
 GTG GGC GAT TAC AGC GTG CTG TAC AAC AGC ACC TTC
 TTC AGC ACC TTC AAG TGC TAC GGC GTG AGC GGC ACC
 AAG CTG AAC GAC CTG TGC TTC AAC AAG GTG TAC GCC
 GAC AGC TTC GTG GTG AAG GGC GAC GAC CTG CGC CAG
 ATC GCC CCC GGC CAG ACC GGC GTG ATC GGC GAT TAC
 AAC TAC AAG CTG CCC GAT GAC TTC ATG GGC TGC GTG
 CTG GCC TGG AAC ACC CGC AAC ATC GAT GCC ACC AGC
 ACC GGC AAC TAC AAC TAC AAG TAC CGC TAC CTG CGC
 CAC GGC AAG CTG GGC CCC TTC GAG CGC GAT ATC AGC
 AAC GTG CCC TTC AGC CCC GAT GGC AAG CCC TGC ACC
 CCC CCC GCC CTC AAC TGT TAC TGG CCC CTC AAC GAT
 TAC GGC TTC TAC ACC ACC ACC GGC ATC GGC TAC CAG
 CCC TAC CGC GTG GTG GTG CTG AGC TTC GAC CTG CTG
 AAC GCC CCC GCC ACC GTG TGC GCC CCC AAG CTG AGC
 ACC GAC CTG ATC AAA AAC CAG TGC GTG AAC TTC AAC
 TTC AAC GGC CTG ACC GGC ACC GGC GTG CTG ACC CCC
 AGC AGC AAC CGC TTC CAG CCC TTC CAG CAG TTC GGC
 GGC GAC GTG AGC GAC TTC ACC GAC AGC GTG CGC GAT
 CCC AAG ACC AGC GAG ATC CTG GAT ATC AGC CCC TGC
 AGC TTC GGC GGC GTG AGC GTG ATC ACC CCC GGC ACC
 AAC GCC AGC AGC GAG GTG GGC CTG TAC CAG GAC
 GTG AAC TGC ACC GAT GTG AGC ACC GCC ATC CAC GCC
 GAT CAG GTG ACC CCC GGC TGG CGC ATC TAC AGC ACC
 GGC AAC AAC GTC TTC CAG ACC CAG GCC GGC TGT CTG
 ATC GGC GGC GAG CAT GTG GAC ACC AGC TAC GAG TGT
 GAT ATC CCC ATC GGC GGC GGC ATC TGC GCC AGC TAC
 CAT ACC GTG AGC CTG CTG GGC AGC ACC AGC CAG AAG
 AGC ATG GTG GCC TAC ACC ATG AGC CTG GGC GCC

[0152] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:6 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:6 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:6, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:6 is shown in Table 11.

TABLE 11

AMINO ACID	Number in SEQ ID NO: 6
A	Ala 43
R	Arg 16
C	Cys 10
G	Gly 30
H	His 5
I	Ile 36
L	Leu 46
K	Lys 25
M	Met 10
F	Phe 28
P	Phe 19
S	Ser 35
T	Thr 38
W	Tyr 4
Y	Tyr 17
V	Val 33
N	Asn 35
D	Asp 26
Q	Gln 34
E	Glu 23

[0153] Using the amino acid composition shown in Table 11, a human codon-optimized coding region which encodes SEQ ID NO:6 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:6 as follows: the 28 phenylalanine codons are TTC, the 46 leucine codons are CTG, the 36 isoleucine codons are ATC, the 10 methionine codons are ATG, the 33 valine codons are GTG, the 35 serine codons are AGC, the 19 proline codons are CCC, the 38 threonine codons are ACC, the 43 alanine codons are GCC, the 17 tyrosine codons are TAC, the 5 histidine codons are CAC, the 34 glutamine codons are CAG, the 35 asparagine codons are AAC, the 25 lysine codons are AAG, the 26 aspartic acid codons are GAC, the 23 glutamic acid codons are GAG, the 10 cysteine codons are TGC, the 4 tryptophan codon is TGG, the 16 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 30 glycine codons are GGC. The codon-optimized coding region designed by this method is presented herein as SEQ ID NO:29.

GAC AGC AGC ATC GCC TAC AGC AAC AAC ACC ATC GCC
ATC CCC ACC AAC TTC AGC ATC AGC ATC ACC ACC GAG
GTG ATG CCC GTG AGC ATG GCC AAG ACC AGC GTG GAC
TGC AAC ATG TAC ATC TGC GGC GAC AGC ACC GAG TGC
GCC AAC CTG CTG CTG CAG TAC GCC AGC TTC TGC ACC
CAG CTG AAC CGG GCC CTG AGC GGC ATC GCC GCC GAG
CAG GAC CGG AAC ACC CGG GAG GTG TTC GAC CAG GTG
AAG CAG ATG TAC AAG ACC CCC ACC CTG AAG TAC TTC
GGC GGC TTC AAC TTC AGC CAG ATC CTG CCC GAC CCC
CTG AAG CCC ACC AAG CGG AGC TTC ATC GAG GAC CTG

-continued

CTG TTC AAC AAG GTG ACC CTG GCC GAC GCC GGC TTC
ATG AAG CAG TAC GGC GAG TGC CTG GGC GAC ATC AAC
GCC CGG GAC CTG ATC TGC GCC CAG AAG TTC AAC GGC
CTG ACC GTG CTG CCC CCC CTG CTG ACC GAC GAC ATG
ATC GGC GCC TAC ACC GGC GCC CTG GTG AGC GGC ACC
GCC ACC GCC GGC TGG ACC TTC GGC GCC GGC GCC GCC
CTG CAG ATC CCC TTC GCC ATG CAG ATG GCC TAC CGG
TTC AAC GGC ATC GGC GTG ACC CAG AAC GTG CTG TAC
GAG AAC CAG AAG CAG ATC GCC AAC CAG TTC AAC AAG
GCC ATC AGC CAG ATC CAG GAG AGC CTG ACC ACC ACC
AGC ACC GCC CTG GGC AAG CTG CAG GAG GTG GTG AAC
CAG AAC CCC GAC GCC CTG AAC ACC CTG GTG AAG CAG
CTG AGC AGC AAC TTC GGC GCC ATC AGC AGC GTG CTG
AAC GAC ATC CTG AGC CGG CTG GAG ATG GTG GAG GCC
GAG GTG CAG ATC GAC CGG CTG ATC ACC GCC CGG CTG
CAG AGC CTG CAG ACC TAC GTG ACC CAG CAG CTG ATC
CGG GCC GCC GAG ATC CGG GCC ACC GCC AAC CTG GCC
GCC ACC AAG ATG AGC GAG TGC CTG GTG CAG AGC
AAG CGG GTG GAC TTC TGC GGC AAC GGC TAC CAC CTG
ATG AGC TTC CCC CAG GCC GCC CCC CAC GGC GTG GTG
TTC CTG CAG GTG ACC TAC GTG CCC AGC CAG GAG CGG
AAC TTC ACC ACC GCC CCC GCC ATC TGC CAG GAG GGC
AAG GCC TAC TTC CCC CGG GAG GGC GTG TTC GTG TTC
AAC GGC ACC AAG TGG TTC ATC ACC CAG CGG AAC TTC
TTC AGC CCC CAG ATC ATC ACC ACC GAC AAC ACC TTC
GTG AGC GGC AAC TGC GAC GTG GTG ATC GGC ATC ATC
AAC AAC ACC GTG TAC GAC CCC CTG CAG CCC GAG CTG
GAC AGC TTC AAG GAG GAG CTG GAC CAG TAC TTC AAG
AAC CAC ACC AGC CCC GAC CTG GAC CTG GGC GAC ATC
AGC GGC ATC AAC GCC AGC GTG GTG AAG ATC CAG AAG
GAG ATC GAC CGG CTG AAC GAG GTG GCC AAG CAG CTG
AAC GAG AGC CTG ATC GAC CTG CAG GAG GAG CTG GGC
TAC GAG CAG TAC ATC AAG TGG CCC TGG

[0154] A codon-optimized coding region encoding SEQ ID NO:56 designed by this method is presented herein as SEQ ID NO:64.

ATG GAC AGC AGC ATC GCC TAC AGC AAC AAC ACC ATC
GCC ATC CCC ACC AAC TTC AGC ATC AGC ATC ACC ACC

-continued

GAG GTG ATG CCC GTG AGC ATG GCC AAG ACC AGC GTG
 GAC TGC AAC ATG TAC ATC TGC GGC GAC AGC ACC GAG
 TGC GCC AAC CTG CTG CTG CAG TAC GGC AGC TTC TGC
 ACC CAG CTG AAC CGG GGC CTG AGC GGC ATC GGC GGC
 GAG CAG GAC CGG AAC ACC CGG GAG GTG TTC GCC CAG
 GTG AAG CAG ATG TAC AAG ACC CCC ACC CTG AAG TAC
 TTC GGC GGC TTC AAC TTC AGC CAG ATC CTG CCC GAC
 CCC CTG AAG CCC ACC AAG CGG AGC TTC ATC GAG GAC
 CTG CTG TTC AAC AAG GTG ACC CTG GCC GAC GCC GGC
 TTC ATG AAG CAG TAC GGC CAG TGC CTG GGC GAC ATC
 AAC GCC CGG GAC CTG ATC TGC GCC CAG AAG TTC AAC
 GGC CTG ACC GTG CTG CCC CCC CTG CTG ACC GAC GAC
 ATG ATC GCC GCC TAC ACC GCC GGC CTG GTG AGC GGC
 ACC GCC ACC GCC GGC TGG ACC TTC GGC GCC GGC GCC
 GGC CTG CAG ATC CCC TTC GGC CAG AAG GGC TAC
 CGG TTC AAC GGC ATC GGC GTG ACC CAG AAC GTG CTG
 TAC GAG AAC CAG AAG CAG ATC GCC AAC CAG TTC AAC
 AAG GCC ATC AGC CAG ATC CAG GAG AGC CTG ACC ACC
 ACC AGC ACC GCC CTG GGC AAG CTG CAG GAG GTG GTG
 AAC CAG AAC GCC CAG GGC CTG AAC ACC CTG GTG AAG
 CAG CTG AGC AGC AAC TTC GGC GCC ATC AGC AGC GTG
 CTG AAC GAC ATC CTG ACC CGG CTG GAC AAG GTG GAG
 GCC GAG GTG CAG ATC GAC CGG CTG ATC ACC GGC CGG
 CTG CAG AGC CTG ACC CAG TAC GTG ACC CAG CAG CTG
 ATC CGG GCC GCC GAG ATC CGG GCC AGC GCC AAC CTG
 GCC GCC ACC AAG ATC ACC GAG TGC ATG CTG GGC CAG
 AGC AAG CGG GTG GAC TTC TGC GGC AAG GGC TAC CAG
 GTG ATG AGC TTC CCC CAG GCC GCC CAC GGC CTG
 GTG TTC CTG CAG CTG ACC TAC GTG CCC AGC CAG GAG
 CGG AAC TTC ACC ACC GCC CCC GCC ATC TGC CAC GAG
 GGC AAG GCC TAC TTC CCC CGG GAG GGC GTG TTC GTG
 TTC AAG GGC ACC AGC TGC TTC ATC ACC CAG CGG AAC
 TTC TTC AGC CCC CAG ATC ATC ACC ACC GAC AAC ACC
 TTC GTG AGC GGC AAC TGC CAG GTG GTG ATC GGC ATC
 ATC AAC AAC ACC GTG TAC GAC CCC CTG CAG CCC GAG
 CTG GAC AGC TTC AAG GAG CAG GTG GAC AAG TAC TTC
 AAG AAC CAC ACC AGC CCC GAC CTG CAG CTG GGC GAC
 ATC AGC GGC ATC AAC GCC AGC GTG GTG AAC ATC CAG

-continued

AAG GAG ATC GAC CGG CTG AAC GAG GTG GCC AAG AAC
 CTG AAC GAG AGC CTG ATC GAC CTG CAG GAG CTG GGC
 AAG TAC GAG CAG TAC ATC AAG TGG CCC TGG

[0155] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:6 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:6 as follows: about 13 of the 28 phenylalanine codons are TTT, and about 15 of the phenylalanine codons are TTC; about 3 of the 46 leucine codons are TTA, about 6 of the leucine codons are TTG, about 6 of the leucine codons are CTT, about 9 of the leucine codons are CTC, about 4 of the leucine codons are CTA, and about 18 of the leucine codons are CTG; about 13 of the 36 isoleucine codons are ATT, about 17 of the isoleucine codons are ATC, and about 6 of the isoleucine codons are ATA; the 10 methionine codons are ATG; about 6 of the 33 valine codons are GTT, about 15 of the valine codons are GTG, about 4 of the valine codons are GTA, and about 8 of the valine codons are GTC; about 6 of the 35 serine codons are TCT, about 8 of the serine codons are TCC, about 5 of the serine codons are TCA, about 2 of the serine codons are TCG, about 6 of the serine codons are AGT, and about 8 of the serine codons are AGC; about 5 of the 19 proline codons are CCT, about 6 of the proline codons are CCC, about 6 of the proline codons are CCA, and about 2 of the proline codons are CCG; about 9 of the 38 threonine codons are ACT, about 14 of the threonine codons are ACC, about 11 of the threonine codons are ACA, and about 4 of the threonine codons are ACG; about 11 of the 43 alanine codons are GCT, about 17 of the alanine codons are GCC, about 10 of the alanine codons are GCA, and about 5 of the alanine codons are GCG; about 7 of the 17 tyrosine codons are TAT and about 10 of the tyrosine codons are TAC; about 2 of the 5 histidine codons are CAT and about 3 of the histidine codons are CAC; about 9 of the 34 glutamine codons are CAA and about 25 of the glutamine codons are CAG; about 16 of the 35 asparagine codons are AAT and about 19 of the asparagine codons are AAC; about 11 of the 25 lysine codons are AAA and about 14 of the lysine codons are AAG; about 12 of the 26 aspartic acid codons are GAT, and about 14 of the aspartic acid codons are GAC; about 10 of the 23 glutamic acid codons are GAA and about 13 of the glutamic acid codons are GAG; about 5 of the 10 cysteine codons are TGT and about 5 of the cysteine codons are TGC; the 4 tryptophan codons are TGG; about 1 of the 16 arginine codons is CGT, about 3 of the arginine codons are CGC, about 2 of the arginine codons are CGA, about 3 of the arginine codons are CGG, about 4 of the arginine codons are AGA, and about 3 of the arginine codons are AGG; and about 5 of the 30 glycine codons are GGT, about 10 of the glycine codons are GGC, about 8 of the glycine codons are GGA, and about 7 of the glycine codons are GGG.

[0156] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must

remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0157] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:6, optimized according to codon usage in humans is presented herein as SEQ ID NO:28.

GAC AGT TCA ATC GCC TAT TCG AAC AAC ACT ATA GCA
ATC CCA ACA AAT TTT TCA ATT TCT ATA ACA ACA GAG
GTG ATG CCA GTG TCC ATG GCA AAG ACT AGC GTA GAC
TGC AAT ATG TAC ATC TGC GGA GAT TCT ACA GAA TGT
GCA AAC TTG CTG CTA CAG TAT GGA TCG TTC TGT ACC
CAG CTC AAC CGG GCG CTG AGC GGC ATT GCT GCC GAA
CAG GAT CGC AAT ACG AGA GAG GTG TTT GCT CAA GTG
AAA CAA ATG TAT AAG ACC CCA ACA TTG AAA TAC TTC
GGT GGA TTC AAT TGC ACT CAG ATT CTG CCA GAC CCA
CTC AAA CCC ACC AAG AGC AGC TTT ATT GAA GAT CTT
CTG TTC AAC AAA GTT ACC TTG GCC GAC GCT GGG TTT
ATG AAG CAA TAC GGT GAG TGC CTG GGC GAC ATT AAC
ACA CGA GAC CTG ATC TGC GCC CAG AAG TTT AAC GGG
CTC ACG GTT TTA CCG CCA CTG CTG ACT GAT GAT ATG
ATT GCC GCT TAC ACT GCG GCC CTT GTG AGT GGT ACC
GCA ACT GCT GCG TGG AGG TTT GGC GCT GGG GCG GCC
TTA CAG ATC CCT TTT GCC ATG CAG ATG GCC TAC AGG
TTC AAT GGA ATT GGT GTC ACT CAG AAT GTC CTG TAC
GAG AAC CAG AAA CAG ATC GCC AAC CAG TTC AAT AAA
GCT ATT TCA CAG ATT CAG GAA TCA CTT ACC ACA ACT
TCC ACG GCA CTC GGT AAA CTG CAG GAC GTG GTG AAT
CAG AAC GCT CAG GCA CTA AAT ACA CTC GTC AAG CAA
CTG AGT TCC AAT TTC GGG GCG ATA TCT AGC GTA TGG
AAC GAC ATC CTC AGT CGG CTC GAC AAA GTG GAG GGC
GAA GTC CAA ATA GAC CTT GAT ATC ACA GGC AGA CTA
CAG TCA TTG CAG CAG TAC GTT ACC CAG CAG TTG ATC
GCG GCC GCT GAG ATA GCA GCG TCC GCC AAT CTG GCG
GCT ACC AAA ATG TCT GAG TGT GTG CTC GGA CAA AGT
AAG CGG GTG GAT TTT TGC GCG AAG GGC TAT CAC CTC
ATC TCC TTC CTT CAA GCA GCA CCC CAG GGA GTC GTT
TTT CTG CAT GTG ACA TAC GTG CTT AGC CAG GAG AGA
AAC TTT ACC ACT GCG CCT GCC ATT TGT CAT GAA GGC
AAA GCT TAT TTT CCC CGC GAG GGG GTG TTC GTT TTC
AAC GGA ACT AGC TGG TTT ATC ACA CAA AGG AAT TTC

-continued

TTC TCC CCC CAG ATC ATC ACC ACC GAC AAC ACC TTT
GTC TCT GGA AAC TGT GAC GTC GTT ATA GGC ATC ATC
AAT AAT ACA GTA TAC GAT CCC CTG CAG CCC GAA CTT
GAC TCT TTC AAG GAG GAA CTA GAT AAG TAC TTC AAG
AAT CAC ACC AGC CGG GAT GTA GAT TTA GGG GAT ATT
AGC GGG ATT AAC GCA TCC GTG GTC AAC ATC CAA AAA
GAG ATT GAC AGA CTG AAC GAA TCG GCG AAG AAC CTG
AAT GAG TCC CTG ATC GAT CTT CAG GAG CTG GGC AAG
TAT GAA CAG TAT ATC AAG TGG CTT TGG

[0158] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:56, optimized according to codon usage in humans is presented herein as SEQ ID NO:65.

ATG GAC AGT TCA ATC GCC TAT TCG AAC AAC ACT ATA
GCA ATC CCA ACA AAT TTT TCA ATT TCT ATA ACA ACA
GAG GTG ATG CCA GTG TCC ATG CAA AAG ACT ACC GTA
GAC TGC AAT ATG TAC ATC TGC GCA GAT TCT ACA GAA
TGT GCA AAC TTG CTG CTA CAG TAT GGA TCG TTC TGT
ACC CAG CTC AAC CGG GCG CTG AGC GGC ATT GCT GCC
GAA CAG GAT CGC AAT ACG AGA GAG GTG TTT GCT CAA
GTG AAA CAA ATG TAT AAG ACC CCA ACA TTG AAA TAC
TTC GGT GGA TTC AAT TTC AGT CAG ATT CTG CCA GAC
CCA CTC AAA CCC ACC AAG AGG AGC TTT ATT GAA GAT
CTT CTG TTC AAC AAA GTT ACC TFG GCG GAC GCT GGG
TTT ATG AAG CAA TAC GGT GAG TGC CTG GGC GAC ATT
AAC CCA GCA GAC CTG ATC TGC GCC CAG AAG TTT AAC
GGG CTC ACG GTT TTA CCG CCA CTG CTG ACT GAT GAT
ATG ATT GCG GCT TAC ACT GCG GCC CTT GTG AGT GGT
ACC GCA ACT GCT GCG TGG ACG TTT GCG CTT AGG GCG
GCC TTA CAG ATC CTT TTT GCC ATG CAG ATG GCC TAC
AGG TTC AAT GGA ATT GGT GTC ACT CAG AAT GTC CTG
TAC GAG AAC CAG AAA CAG ATC GCC AAC CAG TTC AAT
AAA GCT ATT TCA CAG ATT CAG GAA TCA CTT ACC ACA
ACT TCC ACC GCA CTC GGT AAA CTG CAG GAC GTG GTG
AAT CAG AAC GCT CAG GCA CTA AAT ACA CTC GTC AAG
CAA CTG AGT TCC AAT TTC GGG GCG ATA TCT AGC GTA
TTG AAC GAC ATC CTC AGT CGG CTC GAC AAA GTG GAG
GCC GAA GTC CAA ATA GAC GCT CTT ATC ACA GCG AGA

-continued

CTA CAG TCA TTG CAG ACC TAC GTT ACC CAG CAG TTG
 ATC CGC GCC GCT GAG ATA CGA GCC TCC GCC AAT CTG
 GGC GCT ACC AAA ATG TCT GAG TGT GTG CTC GGA CAA
 AGT AAG CGG GTG GAT TTT TGC GGC AAG GGC TAT CAC
 CTC ATG TCC TTC CCT CAA GCA CCC CAC GGA GTC
 GTT TTT CTG CAT GTG ACA TAC GTG CCT AGC CAG GAG
 AGA AAC TTT ACC ACT GCG CCT GCC ATT TGT CAT GAA
 GGC AAA GCT TAT TTT CCC CGC GAG GGG GTG TTC GTT
 TTC AAC GGA ACT AGC TGG TTT ATC ACA CAA AGG AAT
 TTC TTC TCC CCC CAG ATC ATC ACC ACC GAC AAC ACC
 TTT GTC TCT GGA AAC TGT GAC GTC GTT ATA GGC ATC
 ATC AAT AAT ACA GTA TAC GAT CCC CTG CAG CCC GAA
 CTC GAG TCT TTC AAG GAG GAA CTA GAT AAG TAC TTC
 AAG AAT CAC ACC AGC CCG GAT GTA GAT TTA GGG GAT
 ATT AGC GGG ATT AAC GCA TCC GTG GTC AAC ATC CAA
 AAA GAG ATT GAC AGA CTG AAC GAA GTG GCG AAG AAC
 CTG AAT GAG TCC CTG ATC GAT CTT CAG GAG CTG GGC
 AAG TAT GAA CAG TAT ATC AAG TGG CCT TGG

[0159] Another representative codon-optimized coding region encoding SEQ ID NO:6 is presented herein as SEQ ID NO:46.

GAT AGC AGC ATA GCC TAC TCA AAC AAC ACG ATC GCC
 ATC CCC ACA AAC TTT TCC ATT TCC ATA ACT ACC GAG
 GTG ATG CCC GTG AGC ATG GCC AAG ACA TCG GTA GAT
 TGC AAC ATG TAC ATC TGT GGC GAT TCT ACA GAG TGT
 GCC AAC CTG CTG CTG CAG TAC GGC TCT TTC TGC ACG
 CAG CTG AAC AGG GCC TCG TCT GGC ATC GCC GCC GAG
 CAG GAT CGG AAC ACA CGG GAG GTT TTC GCC CAG GTA
 AAG CAG ATG TAT AAG ACG CCC ACT CTG AAG TAC TTC
 GGC GGC TTC AAC TTC TCT CAG ATA CTG CCC GAC CCC
 CTG AAG CCC ACT AAG AAG TCT TTT ATC GAG GAT CTG
 CTG TTC AAC AAG GTT ACC CTG GCC GAT GCC GGC TTT
 ATG AAG CAG TAT GGC GAG TGC CTG GGC GAC ATC AAC
 GCC AGA GAT CTG ATA TGC GCC CAG AAG TTC AAC GGC
 CTG ACT GTG CTG CCC CCC CTG CTG ACT GAC GAC ATG
 ATC GCC GCC TAT ACC GCC CCC GTG GTG AGT GGC ACA
 GCC ACT GCC GGC TGG ACA TTC GGC GCC GGC GCC
 CTG CAG ATC CCC TTC GCC ATG CAG ATG GCC TAC AGA

-continued

TTT AAC GGC ATT GGC GTC ACT CAG AAC GTC CTG TAT
 GAG AAC CAG AAG CAG ATC GCC AAC CAG TTT AAC AAG
 GCC ATA AGC CAG ATC CAG GAG TCA CTG ACA ACG ACA
 AGT ACC GCC CTG GGC AAG CTG CAG GAT GTA GTG AAC
 CAG AAC GCC CAG GCC CTG AAC ACT CTG GTT AAG CAG
 CTG TCT AGC AAC TTC GGC GCC ACT AGT AGT GTT CTG
 AAC GAT ATT CTG TCT AGG CTG GAC AAG GTC GAG GCC
 GAG GTG CAG ATT GAT CGC CTG ATT ACC GGC AGA CTG
 CAG AGT CTG CAG ACT TAT GTA ACT CAG CAG CTG ATC
 AGA GCC GCC GAG ATT CGA GCC TCC GCC AAC CTG GCC
 GCC ACA AAG ATG TCT GAG TGC CTG GGC CAG AGT
 AAG AGG GTT GAC TTC TGC GGC AAG GGC TAT CAT CTG
 ATG TCT TTT CCC CAG GCC GCC CCC CAG GGC GTC GTG
 TTC CTG CAC GTA ACT TAC GTG CCC AGT CAG GAG AGA
 AAC TTT ACC ACT GCC CCC GCC ATC TGC CAG GAG GGC
 AAG GCC TAC TTC CCC AGA GAG GGC GTG TTT GTG TTC
 AAC GGC ACA TCT TGG TTC ATC ACC CAG AGG AAC TTT
 TAC AGC CCC CAG ATC ATA ACA ACT GAC AAC ACT TTC
 GTT TCG GGC AAC TGC GAC GTA GTG ATC GGC ATA ATA
 AAC AAC ACC TTG TAC GAT CCC CTG CAG CCC GAG CTG
 GAC AGC TTT AAG GAG GAG CTG CAG AAG TAC TTT AAG
 AAC CAT ACC TCA CCC GAT GTG GAC CTG GCC GAC ATT
 TCT GGC ATA AAC GCC TCC GTC GTC AAC ATC CAG AAG
 GAG ATA GAT AGA CTG AAC GAG GTT GCG AAG AAC CTG
 AAC GAG TCC CTG ATC GAT CTG CAG GAG CTG GGC AAG
 TAC CAG CAG TAT ATA AAG TGG CCC TGG

[0160] Another representative codon-optimized coding region encoding SEQ ID NO:56 is presented herein as SEQ ID NO:66.

ATG GAT AGC AGC ATA GCC TAC TCA AAC AAC ACG ATC
 GCC ATC CCC ACA AAC TTT TCC ATT TCC ATA ACT ACC
 GAG GTG ATG CCC GTG AGC ATG GCC AAG ACA TCG GTA
 GAT TGC AAC ATG TAC ATC TGT GGC GAT TCT ACA GAG
 TGT GCC AAC CTG CTG CTG CAG TAC GGC TCT TTC TGC
 ACG CAG CTG AAC AGG GCC CTG TCT GGC ATC GCC GCC
 GAG CAG GAT CGG AAC ACA CGG GAG GTT TTC GCC CAG
 GTA AAG CAG ATG TAT AAG ACG CCC ACT CTG AAG TAC
 TTC GGC GCC TTC AAC TTC TCT CAG ATA CTG CCC GAC

-continued

```

CCC CTG AAG CCC      ACT AAG AGG TCT TTT ATC GAG GAT
CTG CTG TTC AAC AAG GTT ACC CTG GCC GAT GGC GGC
TTT ATG AAG CAG TAT GGC GAG TGC CTG GGC GAC ATC
AAC GCC AGA GAT CTG ATA TGC GCC GCG AAG TTC AAC
GGC CTG ACT GTG CTG CCC CTG CTG ACT GAC GAC
ATG ATC GCC GCC TAT ACC GCC GGC CTG GTG AGT GGC
ACA GCC ACT GGC GGC TGG AGA TTC GGC GGC GGC GGC
GCC CTG CAG ATC CCC TTC GGC ATC CAG ATG GCC TAC
AGA TTT AAC GGC ATT GGC GTC ACT CAG AAC GTC CTG
TAT GAG AAC CAG AAG CAG ATC GCC AAC CAG TTT AAC
AAG GCC ATA AGC CAG ATC CAG GAG TCA CTG ACA ACG
ACA AGT ACC GCC CTG GGC AAG CTG CAG GAT GTA GTG
AAC CAG AAC GGC CAG CTG AAC ACT CTG GTT AAG
CAG CTG TCT AGC AAC TTC GGC GCC ATC AGT AGT GTT
CTG AAC GAT ATT CTG TCT AGG CTG GAC AAG GTC GAG
GCC GAG GTG CAG ATT GAT GGC CTG ATT ACC GGC AGA
CTG CAG AGT CTG CAG ACT TAT GTA ACT CAG CAG CTG
ATC AGA GGC GGC GAG ATT GGA GGC TCC GGC AAC CTG
GCC GGC ACA AAG ATG TCT GAG TGC GTC CTG GGC CAG
AGT AAG AGG GTT GAC TTC TGC GGC AAG GGC TAT CAT
CTG ATG TCT TTT CCG CAG GCC GGC CCC CAC GGC GTC
GTG TTC CTG CAC GTA ACT TAC CTG CCC AGT CAG GAG
AGA AAC TTT ACC ACT GGC CCC GGC ATC TGC CAC GAG
GGC AAG GCC TAC TTC CCC AGA GAG GGC GTG TTT GTG
TTC AAC GGC ACA TCT TGG TTC ATC ACC CAG AGG AAC
TTT TTC AGC CCC CAG ATC ATA ACA ACT GAC AAC ACT
TTC GTT TCG GGC AAC TGC GCA GTA GTG ATC GGC ATA
ATA AAC AAC ACC GTG TAC GAT CCC CTG CAG CCC GAG
CTG GAC AGC TTT AAG GAG GAG CTG GAC AAG TAC TTT
AAG AAC CAT ACC TCA CCC GAT GTG GAC CTG GGC GAC
ATT TCT GGC ATA AAC GGC TCC GTC GTC AAC ATC CAG
AAG GAG ATA GAT AGA CTG AAG GAG GTT GCC AAG AAC
CTG AAC GAG TCC CTG ATC GAT CTG CAG GAG CTG GGC
AAG TAC GAG CAG TAT ATA AAG TGC CCC TGG

```

[0161] In certain embodiments, a codon-optimized coding region encoding the full-length SARS-CoV spike protein (SEQ ID NO:23) is optimized according to any plant, animal, or microbial species, including humans. A codon-optimized coding region encoding SEQ ID NO:23 was first established using the "uniform" optimization protocol described above. However, certain additional adjustments to the sequence were carried out in order to eliminate, for

example, newly opened reading frames being created on the opposite strand, splice acceptors, stretches of identical bases, or unwanted restriction enzyme sites. Making such adjustments is well within the capabilities of a person of ordinary skill in the art.

[0162] A codon-optimized coding region encoding SEQ ID NO:23 is conveniently synthesized as smaller fragments, which are then spliced together using restriction enzyme sites engineered into the sequence fragments. Examples of fragments of codon-optimized coding regions encoding SEQ ID NO:23 are as follows.

[0163] SEQ ID NO:57 has the following sequence:

```

GTGCACATGGTTATCTTCTGCTGCTTCTCTCAACCTCACCAGCGCGAAGCA
TCTGTGATAGGTGCAACACCTTTCAGGACGTGACGAGGCCCACTACACCC
AGCACACCGACGATGAGGGCGGTACTACCCCGAGAGATTTTCAGA
AGGACACCCCTGTACCTCACCACGAGACCTGTTCCTGCCCTCTCAGCA
CTGACACCGGCTTCACACACATCAACACACCTTCGGCAACCCCTGTATCC
CTTTCAAGGACGGCATCTACTTCGCCGCCACGAGAGAGCAATGTGGTG
CGGGGTGGGTGTTCGGGACGACCATGAACAAGAGCCAGAGCGGTGAT
CATCATCAACAACAGCACCAAGGTGGTGATCCGGGCTGCAATTTGAGCG
TGTGGACACACCTTTCTTGGCGGTGTCCAAACCTATGGGACACCGAGCC
CACACCATGATCTTCGACAAACGCTTCAACTGCACCTTGAGTACATCAG
CGACGCTTCAGCGGTGATGTGAGCGAGAGAGCGGCACTTCAAGGACCC
TGGCGAGGTGCTGTTCAGAACAAAGGACGGCTCTGTGACGTGTACAG
GGCTACACGCCATCGAGGTGTGAGAGACCTGCCGAGCGGTTCACAC
CTGAAAGCCATCTTCAAGCTGCCCGTGCGATCAACATCAACCACTTCC
GGGCGATCTCAGCGCTTACGCTGCCCGAGATCTCGGGGACACGAGC
GGCGCTGCTACTTCTGTGGCTACTGTGAAGCTACCACTGCTTCATGTGAA
GTACAGCAGAGAGCGCACCATCAAGGATCGGCTGAGCTGCAGCGAGAAC
CCCTGGCGGAGCTGAAGTGACAGGTGAAGAGCTTCGAGATCGACAAGGCG
ATCTACGACGACGCAACTGTAGAGGTGGTGTACGGCGGATGTGTGTAG
GTTCGCCAATFACACCAACCTGTGCCCTTCGGCGAGGTGTTCAAGGCCA
CGAGTGTCCCTAGCGGTGAGCGCTGGGAGGCGAAGAGATCAACCACTGCG
GTGGCCGATTACAGCGGTGCTGTAACACTCCACTTCTTCAGCACTTCAA
GTGCTACGGGGTGAGCGCCACCAAGCTGAAGCACTGCTGTGCTTCAGCAAG
TGTACGCGACATCACTGTGTGTGAAGGCGGACGACGATGATGCGCC
CTTGCCGACACCGGGGTGATCGCGGACTACAACCTCAAGCTT

```

[0164] Nucleotides 7 to 1242 of SEQ ID NO:57 encode amino acids 1 to 412 of SEQ ID NO:23, with the exception that amino acid 2 (Phenylalanine, (F)) of SEQ ID NO:23 is replaced with valine (V). The translation product of nucleotides 7 to 1242 of SEQ ID NO:57 is presented herein as SEQ ID NO:58.

MVIFLLFLTLTSGSLDDBCTTFDDVQALPNYTHQTSMSRGVYYIPDRFBS
 DTLYLTDQLFLPFYSNVTGPHITNHPGPNVFFPKDGIYFAATKESNVVR
 GWVFGSTMNKNSQSVIIINNSHNVRIRACNFELCDNPFVAVSKPMGTQTR
 THIFDPAFCTPEYIESDAFLDVSKEGPNFLHREFPVFNKDGFLYVYKG
 YQPIDVVRDLPSGFNTLKLPIFLPLGGINITNFRAILTAFSPAQDIWOTS
 AAAYFVGYLKFTFPMKLIDENQITDAVDCSQPLAEKLSVKFSFIDEG
 IYQTSNFRVPSGQVVRFFHITNLCLFFGEVFNATKFPSTYANERKKIENC
 VADTSVLYNSTFTTFKCYQVSATKLNCLCFSHVYAGDSFVVGDDVRGIA
 PQGTQVIADNYTKL

[0165] Nucleotides 1 to 6 of SEQ ID NO:57, GTCGAC, is a recognition site for the restriction enzyme Sal I. Nucleotides 1237 to 1242 of SEQ ID NO:57, AAGCTT, is a recognition site for the restriction enzyme Hind III.

[0166] SEQ ID NO:59 has the following sequence:

AAGCTTCCCGAGACTTCATGGGCTGCTGGCTGGAACACGAAA
 CATCGACGCCACCTCCACCGCACTACAAATACAACTACCGCTACCTGA
 GGCCACGCAAGCTGAGACCTTCGAGCGGAGCATCTCCAACTGCCCTTC
 AGCCAGCGCAAGCCCTCAGCCGCCCTGCCCTGCCCTGAATGCTACTGGCC
 CCTGAACGACTACGGCTTCTACACGACACCGGATCGGCTATCAGCCCT
 ACAGAGTGTGTGTCTGAGCTTGCAGCTCTGGAACGCCCTTGCACCGCTG
 TGCGGCCCCAGCTGAGCACCGACCTCATCAAGAACAGTGTGGTGAACCT
 CATCTTCAACGGCTTACCGGCAACCGGCTGCTCAGCCCGCAGCAAGA
 GATTCAGCCCTTCCAGCACTTCGCGAGGAGCTGAGCGATTTCAACGAC
 AGCGTAGGGATCTTAGAACGAGGAGATCTCGACATCAGCCCTTCGAG
 CTCGCGCGCTGTCCGTATCACCCTCGGCAACCAACCGCAGCGAGG
 TGCGCGTGTCTTACGAGGCTGGAATCGACCGAGCTGAGCAGCGCATC
 CAGCGGACCGCTCAGCCCGCTGAGAGCTCTAGAGCTCAGCGACCGCAACA
 CGTGCTCCAGACCGGCGGCTGCTCATCGCGCGGAGCACGTGGACA
 CCAGCTCAGTGTGACATCCCATGAGGAGCGGCTCTGCCACGCTAC
 CACACCGCTGAGCTGCTGAGAGCAGCAGCGAGGAGGATCTGGCCCTA
 CAGCATGCTGTGGCGCGGAGCAGGATCATGAGCTCAGCAACCAACCA
 TCGCATCCCCACCAATCTAGCATCTCATCAGCAGCGAGTGTATGCC
 GTGAGCATGCCCAAGCAGCGCTGATGCAACATGTACATCTCGGCGCA
 CAGCAGGAGTGGCCCACTGCTGCTGCAATGAGCAGCTCTGCACCC
 AGCTGACAGAGCCCTGAGCGGATTCGCGCGGAGCAGGACAAACACC
 AGGAGGTTCTCGCCAGGTGAAGCAGATGTATAAGACCCCAACCTGAA
 GTACTTGGCGGTTTCAACTTCAGCCAGATCTGCCGATCTCTGAAGC
 CCACCAAGCGAGCTTCTATCGAGGAGCTGCTTTCAAGAGGTGACCGCTG
 GCGGACCGCGGCTTTATGAAGCACTAGCGCGAGTGGCTGGCGATATCAA

-continued

CGCCAGGAGCTCATCTGCGGCCAGAGATTCACGGCTTGACCTGCTGCTG
 CCGCTGCTGCTCAGGATGATATGATCGCGCTATACAGCGCCGCTGTGTG
 TCAGGCAACGCCACCGCGCGCTGAGCTTTGCGCGCGGAGCGCCCTGCA
 GATCCCTTGCCTCAGCATGAGCTGGCTACCGGT

[0167] Nucleotides 1 to 1431 of SEQ ID NO:59 encode amino acids 411 to 887 of SEQ ID NO:23. Nucleotides 1 to 6 of SEQ ID NO:59, AAGCTT, is a recognition site for the restriction enzyme Hind III. Nucleotides 1237 to 1242 of SEQ ID NO:59, ACCGGT, is a recognition site for the restriction enzymes Age I and PnaI.

[0168] SEQ ID NO:60 has the following sequence:

ACCGGTCAATGGCATCGGCTGACCCAGACAGTGGTGTACGAGAACGAG
 AAGCAGATGCCAACAGTTCATTAAGGCCATCTCCAGATCAGGAGAG
 CCTCACCCACACAGCACCGCTGCGGCAAGCTCAGGACGCTGTGAACCC
 AGAACCCGAGCGCTGAAATCCCTGCTGAACGACTGACGAGCAACTTC
 GGCGCATCAGCAGGCTGTGAAGCATCTGACGCGGCTGTGATAGGT
 GGAGCGGAGGTGCAGATGACAGCATCATCCCGCCAGCATCAGAGGCC
 TGCAGCTCATCTGACCCAGCGCTCATCAGAGCGCGGAGATCAGAGCC
 AGCGCATCTGCGCGCACAGATGAGGAGTGTGCTGCTGGCGCAGAG
 CAGGAGAGTGGATCTCTCGGCGAAGGCTATCATCTCATGAGCTTCCCTC
 AGGCGCTGCCCGCGGCTGTGTCTCTGCACGTGAGCTGTGCTGCTGAGC
 CAGGAGAGGAATTTACACCGCCCGCCAGCTCTGCCAGGCGCAAGGC
 CTACTTCCCGAGAGAGGCGCTGTGTGTGTTTAAAGCCACCGACTGTGTTCA
 TCACCGAGCGGAATCTTTCAGCGCCGAGATCATCAGCAGCAACACC
 TGTGTGTGCGCAATGTGACGTGTGATCGGCTATCATTAATACAGC
 TGTACGACCCCTGAGCGCGAGCTGATGATCTCAAGAGAGAGCTGAGC
 AAGTACTTCAAGAACCAACTTCCCGGAGGTGTGAGCTGGCGCATCAG
 CGCATCATGCTGAGTGTGAGCATCTCAGAGAGGATGTACCGGCTGA
 ACGAGTGGCCAGAACTGTGACGAGAGCGCTCATGACCTCAGGAGGT
 GGAAAGTACGAGCAGTACATCAAGTGGCGCTGTGACGTGTGTGCTGCGCT
 CATCTGCGCGCTCATGCGCATCTGTATGTGACCATCTGCTGTGTGCA
 TGACGAGCTGTCTGCTCTGCTGAGGAGCGCTGACAGTGTGGCAGCTGC
 TGCAGTGTGACGAGGAGCAGGAGCGCCGTGCTGAGGCGGTGAGACT
 GCATCACTTGAAGATCT

[0169] Nucleotides 3 to 1109 of SEQ ID NO:60 encode amino acids 887 to 1255 of SEQ ID NO:23. Nucleotides 1 to 6 of SEQ ID NO:60, ACCGGT, is a recognition site for the restriction enzymes Age I and PnaI. Nucleotides 1113 to 1118 of SEQ ID NO:59, AGATCT, is a recognition site for the restriction enzyme Bgl II.

[0170] SEQ ID NOs 57, 59, and 60 are then spliced together using the restriction enzyme sites described above

to produce a codon-optimized coding region encoding SEQ ID NO:23 in its entirety, with the exception that amino acid 2 (Phenylalanine, (F)) of SEQ ID NO:23 is replaced with valine (V). The spliced sequence is presented herein as SEQ ID NO:61.

GTGCATGGTTATCTTCTGCTGTCTCTACCTCCAGCGGAGCGA
TCTGNTAGGTGCACCACTTGGAGGAGCTGCAAGGCCCACTACACCC
AGCAGACCCAGCAGCATGAGGGCGTGTACTACCCGAGGATTTTCAGA
AGCGACACCTGTACTCACCAGGAGCTGTTCCTGCCCTTCTACAGCAA
CTGTACCGCTTCCAGACCATCAACGACACTCTCGGCAACCCGTGATCC
CTTTCAAGAGCGCATCTACTCTCGCCGCCACCGAAGAGCAATGTGTGT
CGGGCTGGGTGTTTCCGAGCACCATTGAACCAAGAGCCAGAGCGTGAT
CATCATCAACCAACGACCAACAGTGTGTATCCGGCCCTGCAATTTGAGC
TGTGTGCAAACTCTTCTGCGCGTGTCCAAACCTTATGGCACCAGACC
CAACACATGATCTTGACAGACGCTCTCAACTGACCTTCGATACATCAG
CGACGCTCTCAGCGTGATGTGAGCGAAGAGGCGCACTTCAAGACC
TGCGGAGTTGTGTTCAGAAACAGAGAGCGCTTCTGTACGTATACAG
GGCTACAGCCCATCGAGCTGTGAGAGACCTGCCGAGCGCTTCAACAC
CTGAAGCGCATCTCAAGCTGCCCTTGGGATCAACATCAACCACTTCC
GGCCATCTCAACGCTTTAGCCTTGCCAGGATCTTGGGGACCAAGC
GCGCGTGCCTACTTGTGGCGTACCTGAAGCTTACCACCTTCACTGCTGAA
GTACAGCAGAAACCGCACCATTACCGATGCCGTGAGCTCCAGCAGAAC
CCCTGGCCGAGCTGAAGTGCAGCTGTGAAGCTTGTAGATCGACAGAGGC
ATTCACCAAGACCACTTCTAGAGTGTGCTACCGCGGATGTGTGTGAG
GTCTCCCAATATCAACAACTGTGCGCCCTTGGCGAGGTGTTCAAGCCA
CCAAAGTTCCCTAGCGTGTAGCGCTGGGAGCGAAGAGATCAGCAACTGC
GTGGCGGATTACAGCGTCTGTCAACTCCACTTCTTCAAGCACTTCGA
GTGTACGCGTGTGAGCGCCCAAGCTGAGACGACTGTGCTTCAAGAAC
TGTAAGCGACTCAATGTGTGAGGAGGAGCGAGCTGTGAGCAGATGCC
CTGTGCGAGACCGCGGTGATCGCGACTTCAACTCAAGCTTCCGAGGA
CTTCATGGCTGTGCTGTGCTGCTGAACACGAGAACTGAGGCCCACT
CCACCGCAACTACAACTTACAGTACCGCTACTTGAGGCGAGCGCAGTGS
AGACCTTCTGAGCGGAGCATCTCAACGTGCCCTTCAAGCCCGAGCGCAA
GCCCTGCAACCCCGCTGCCCTGACTGCTACTTGCGCCCTGAGACGCTACG
GCTCTTACACCAACCGGATCGGCTATGACGCTTACAGGCTAGAGTGTGTG
CTGAGCTTCTGAGTCTGCTGAGCGCCCTGCCACGCTGTGCGGCCCAAGCT
GAGCAACGACTCATCAGAAACAGTGTGCTGACTTCAACTTAAAGGCC
TCACCGGACCGCGCTGCTCAACCCGAGCAGCAAGATTCAGGCCCTTCT
GAGCATGTCGCGAGGAGCTGTGAGAGATTTCAACGACAGCGTGAAGGATCC
TAAGACCAAGCACTCTGACATCAACGCTTCAGACTTCGCGCGCGTGT

-continued

CCGTGATCACCCCGGACCAACGCCAGCAGGAGGTGCCGTGCTATC
CAGGACGTGAACCTGCACCGAGCTGAGCAGCCGCTACACCGCGACAGCT
CACCCCCGCTGGAGATCTACAGCACCAGCAACAGCTGTTCAGAGCC
AGGCCGCTGCTCTATGCGCGCGAGCATGTGAGACACAGCTAGAGTGT
GATCTCCCATCGAGCGCGCATCTGCGCGAGCTACACACCGTGAAGCT
GTGAGAGACACCGCAGAGAGAGCATGTGCGCTACACCAAGAGCTGTG
GGCCGAGCAGCAGCATCGCTACAGCAACACACATCGCATCCCGACCC
AACTTCAGCATCTCCATCAACACGAGGTGATGCCGTGAGCATGGCAA
GACACAGGTGAGATTGCAACATGTATCTCTGCGGCGACAGCAACCGAGTGC
CCAACCTGCTGTGCAATGAGCGAGCTTCTGCGACCGCTGAACAGAGCC
CTGAGCGGCTATTCGCGCGAGCAGGACAGAAACACAGGAGGTGTGTGCG
CGAGTGAAGCAGATGTATAAGACCCCGACCTGAAGTACTTCGCGCGGT
TCAACTTCAGCGAGATCTCTGCGCGTCTCTGAAGCCCGCAAGCGAGGCG
TTCATCGAGGAGCTGTCTGTCAACAGAGTGCACCTGGCGCAAGCCCGCTT
TATGAAGCAGTACGCGGAGTGTCTGGCGATATAACCGCAGAGCTCTCA
TCTGCGCCAGAGATTCAACGCTTGACGCTGCTGCCCTCTGCTCAAC
GATGATATGATGCGCGCTATACAGCGCCGCTGTGTGAGCGACCGCCAC
CGCGCGCTGGACCTTTGGCGCGGAGCGCGCTGCAAGTCCCTTCGCGCA
TCAGATGGCTCACTGAGCTTCAATGGCATGGCGTAGCCAGAACTGCTG
TAAGAGAACAGAGCAGATCTGCCAACAGTTCAATAGGCGATCTGCCA
GATCAAGAGAGCTCCACACCAACAGCAGCCGCTGGCGAGCTGAGG
ACGTGTGAACAGAGAGCGCGGCTCAATACCTGTGTGAAGCAGCTG
AGCAGCACTTCGCGCGCATCAGCAGCTGTGTGAACGATCTTGCAGAG
CTGTGATAGGTGAGGCGGAGGTGTGAGATCGACAGACTCACTCAACGCGA
GACTGCAGAGCTTGCAGACTCACTGACCGAGCAGTCACTCAGAGCGCGC
GAGATCAGAGCCAGCGCAATCTGCCGCCACCAAGATGAGCGAGTGTGCT
CTGTGGCCAGAGCAAGAGATGGACTTCTGCGCGAAGGCTCACTCACTCA
TGAGCTTCCCTCAGCGCCCTCCCGACGGCTGTGTGCTTCTGACATGACC
TACTGTCTGAGCAGGAGGAGAAATTCACACCGCGCCAGCACTGTGCA
CGAGGCAAGGCTTACTTCCCGAGAGGGCGGTGTGTGTTTAAAGCGA
CGAGCTGCTTCACTACCCAGCGGAATCTTCAAGCCCGAGTCACTCAC
ACAGACAACACTTGTGTGCGCAATTCGAGAGTGTGATCGGCATCAT
CAATAAACGCTGTGAGAGCCCTTCGACCGCGAGCTGTGATAGCTCAAGG
AGAGCTTGAAGTACTTCAAGAACCACTCTCCCGAGCTGGAGCTGT
GGCGCATACGCGGCTCAATGCGCGAGCTGTGTGAGATCTCAGAGAGGAT
CGACCGGCTGAACGAGGTGCGCAAGACCTCAACGAGAGGCTCATCGACC
TGCAAGGCTGGAAAGTACAGAGGATACATCAAGTGGCGTGTGATGCTG
TGGCTGGCTTCTATGCGCGGCTCATGCGCATCTGTATGTGAGCATCTCT
GCTGTGCTGATGACCGAGCTGCTGCTGCTGTGAAGGGCGCTGACACT

-continued

GTGGCAGCTGCTGCAGTTCGACGAGGAGCACTCAGAGCCGCTGCTGAAG

GGCGTGAACTGCTACCTACCTGGAAGATCT

[0171] The translation product of nucleotides 7 to 3771 of SEQ ID NO:61 is presented herein as SEQ ID NO:62

HVFLLFLTLTSGSDLRCTTFDDVQAPNYQTSTSSNRGVYYPDEIFRSD

TVLTQDLFLFTYSVHGFTHNHTFGNVPVFPKDGIIYFAATEKSNVVRG

LVFTSTNRKSKQSIVIIINNSTNVVRACNFELCDNFFFAVKNGTQTHTN

IFDRAPKCTFEYISDAFLDVSEKSGNFKHLEFVFNKMDGLFVYKGYQ

PIDVVRDLPSGFNFKIPFLPLGINITHFAILTAFSQAQDIMGTSAAA

YFVGYLKPTTPLMLYDENSTTDAVDCSNFLAELKCSVKSEFIDKGIYQ

TSNFRVPSGSDVVRFTNITLCPFGVFNATKFPFSVYMERKKISNCHVAD

YSVLNSTFTTFFSTFACVGSATKLNLDLCPNHYADSPVVGDDVVRQIAPGQ

TGVIAIDYNYKLDDPMGCVLAANNTRINDATSTGHNYHYKYLRLHKLKRF

ERDISHVFSPDGKPCFTPALCNYFLNDYGFYTTTGIGYQFVRVVLFSF

ELLNAPATVCGPKSLTDLINQCNVFNFMGLTGVLTPSCKRFPQFQCF

GRDVSDFTSVDRPKTSEILDSPCGSGSVITPOTNASSVAVLYQDV

NCTDVSTAIHAQPLTPANRIYSTGNVFPQTQACLLIGAHVDTSEYCDIF

IGAGTCASVHTVSLRSTSKSIVATYMSLGADSSIAYSNNTIAIPNFS

ISITTEVHPVSMARSTVDCNRIYCGDSTECANILLQYGSFCTQLNRALSG

IAAEQDNTREVFQAQKMYKPTPLKYFGGPNFSLDEPLKPTKRSFTE

DLLFNKVTADMGFNKQYGCIGDINARLCAQNFNGLTVLFLTYDM

IAAYTAALVSGTATAGWTGAGAAQLIPFAMQMYRPFNGIGVTCNVLIN

QGAIQINQFKNAISQISLTTSTALQKLDVVMKRAQMLNTVLKGLSH

FGAISVLDLILSRDLKVEASVQIDRLITGRQLSLQTVTVTQQLRAAIR

ASANLAANTKMSCEVLGQSKSRVFCGQYKLMSPFQAAPRGVFLHVTYVP

SQGRNFTAPALCHEGKAYTFREGVVFNSTGFWPTIQNFVSPQIITDND

TFVSGNCDVIGIINNTVDFPLQELDSFKRELKQPKMRTSPDVLGDII

SGINASVVMIQEIDRLNEVAKNINELSLDLQLGKRYEQYIKNFWYVWLG

FIAGLIALIVHVTLLCCMTSCCSCLNGKACSGSCCKFDEDDSFVLKGV

KLHNT

[0172] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:8 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:8 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:8, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:8 is shown in Table 12.

TABLE 12

AMINO ACID	Number in SEQ ID NO: 8	
A	Ala	84
R	Arg	41
C	Cys	33
G	Gly	77
H	His	14
I	Ile	73
L	Leu	92
K	Lys	57
M	Met	19
F	Phe	79
P	Pro	57
S	Ser	93
T	Thr	94
W	Tyr	10
Y	Tyr	52
V	Val	89
N	Asn	81
D	Asp	71
Q	Gln	55
E	Glu	40

[0173] Using the amino acid composition shown in Table 12, a human codon-optimized coding region which encodes SEQ ID NO:8 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:8 as follows: the 79 phenylalanine codons are TTC, the 92 leucine codons are CTG, the 73 isoleucine codons are ATC, the 19 methionine codons are ATG, the 89 valine codons are GTG, the 93 serine codons are AGC, the 57 proline codons are CCC, the 94 threonine codons are ACC, the 84 alanine codons are GCC, the 52 tyrosine codons are TAC, the 14 histidine codons are CAC, the 55 glutamine codons are CAG, the 81 asparagine codons are AAC, the 57 lysine codons are AAG, the 71 aspartic acid codons are GAC, the 40 glutamic acid codons are GAG, the 33 cysteine codons are TGC, the 10 tryptophan codon is TGG, the 41 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 77 glycine codons are GGC. The codon-optimized coding region designed by this method is presented herein as SEQ ID NO:31.

ATG GAC GCC ATG AAG CGG GGC CTG TGC TGC GTG CTG
CTG CTG TGC GGC GGC GTG TTC GTG AGC CCC AGC GCC
CGG GGC AGC GGC AGC GAC CTG GAC CGG TGC ACC ACC
TTC GAC GAC GTG CAG GCC CCC AAC TAC ACC CAG CAC
ACC AGC AGC ATG CGG GGC GTG TAC TAC CCC GAG GAG
ATC TTC CGG AGC GAC ACC CTG TAC TAC ACC CAG GAC
CTC CTC GTG CCC TTC GAC AAC CTG ACC GGC TTC
CAC ACC ATC AAC CAC ACC TTC GGC AAC CCC GTG ATC
CCC TTC AAG GAC GGC ATC TAC TTC GCC GGC ACC GAG
AAG AGC AAC GTG GTG CGG GGC TGG GTG TTC GGC AGC

-continued

ACC ATG AAC AAC AAG AGC CAG AGC GTG ATC ATC ATC
 AAC AAC AGC ACC AAC GTG GTG ATC CGG GGC TGC AAC
 TTC GAG CTG TGC GAC AAC CCC TTC TCC GCC GTG AGC
 AAG CCC ATG GGC ACC CAG ACC CAC ACC ATG ATC TTC
 GAC AAC GGC TTC AAC TGC ACC TTC GAG TAC ATC AGC
 GAC GGC TTC AGC CTG GAC GTG AGC GAG AAG AGC GGC
 AAC TTC AAG CAC CTG CGG GAG TTC GTG TTC AAG AAC
 AAC GAC GGC TTC CTG TAC GTG TAC AAG GGC TAC CAG
 CCC ATC GAC GTG GTG CGG GAC CTG CCC AGC GGC TTC
 AAC ACC CTG AAG CCC ATC TTC AAG CTG CCC CTG GGC
 ATC AAC ATC ACC AAC TTC CGG GGC ATC CTG ACC GGC
 TTC AGC CCC GGC CAG GAC ATC TGG GGC ACC AGC GGC
 GGC GGC TAC TTC GTG GGC TAC CTG AAG CCC ACC ACC
 TTC ATG CTG AAG TAC GAC GAG AAC GGC ACC ATC ACC
 GAC GGC GTG GAC TGC AGC CAG AAC CCC CTG GGC GAG
 CTG AAG TGC AGC GTG AAG AGC TTC GAG ATC GAC AAG
 GGC ATC TAC CAG ACC AGC AAC TTC CGG GTG GTG CCC
 AGC GGC GAC GTG GTG CGG TTC CCC AAC ATC ACC AAC
 CTG TGC CCC TTC GGC GAG GTG TTC AAC GGC ACC AAG
 TTC CCC AGC GTG TAC GGC TGG GAG CGG AAG AAG ATC
 AGC AAC TGC GTG GGC GAC TAC AGC GTG CTG TAC AAC
 AGC ACC TTC TTC AGC ACC TTC AAG TGC TAC GGC GTG
 AGC GGC ACC AAG CTG AAC GAC CTG TGC TTC AGC AAC
 GTG TAC GGC GAC AGC TTC GTG GTG AAG GGC GAC GAC
 GTG CGG CAG ATC GGC CCC GGC CAG ACC GGC GTG ATC
 GGC GAC TAC AAC TAC AAG CTG CCC GAC GAC TTC ATG
 GGC TGC GTG CTG GGC TGG AAC ACC CGG AAC ATC GAC
 GGC ACC AGC ACC GGC AAC TAC AAC TAC AAG TAC CGG
 TAC CTG CGG CAC GGC AAG CTG CGG CCC TTC GAG CGG
 GAC ATC AGC AAC GTG CCC TTC AGC CCC GAC GGC AAG
 CCC TGC ACC CCC CCC GGC CTG AAC TGC TAC TGC CCC
 CTG AAG GAC TAC GGC TTC TAC ACC ACC ACC GGC ATC
 GGC TAC CAG CCC TAC CGG GTG GTG GTG CTG AGC TTC
 GAG CTG CTG AAC GGC CCC GGC ACC GTG TGC GGC CCC
 AAG CTG AGC ACC GAC CTG ATC AAG AAC CAG TGC GTG
 AAC TTC AAC TTC AAC GGC CTG ACC GGC ACC GGC GTG
 CTG ACC CCC AGC AAG CAG CGG TTC CAG CCC TTC CAG
 CAG TTC GGC CGG GAC GTG AGC GAC TTC ACC GAC AGC

-continued

GTG CGG GAC CCC AAG ACC AGC GAG ATC CTG GAC ATC
 AGC CCC TGC AGC TTC GGC GGC GTG AGC GTG ATC ACC
 CCC GGC ACC AAC GGC AGC AGC GAG GTG GGC GTG CTG
 TAC CAG GAC GTG AAC TGC ACC GAC GTG AGC ACC GGC
 ATC CAC GGC GAC CAG CTG ACC CCC GGC TGG CGG ATC
 TAC AGC ACC GGC AAC AAC GTG TTC CAG ACC CAG GGC
 GGC TGC CTG ATC GGC GGC GAG CAG GTG GAC ACC AGC
 TAC GAG TGC GAC ATC CCC ATC GGC GGC GGC ATC TGC
 GGC AGC TAC CAC ACC GTG AGC CTG CTG CGG AGC ACC
 AGC CAG AAG AGC ATC GTG GGC TAC ACC ATG AGC CTG
 GGC GGC GAC AGC AGC ATC GGC TAC AGC AAC AAC ACC
 ATC GGC ATC CCC ACC AAC TTC AGC ATC AGC ATC ACC
 ACC GAG GTG ATG CCC GTG AGC ATG GGC AAG ACC AGC
 GTG GAC TGC AAC ATG TAC ATC TGC GGC GAC AGC ACC
 GAG TGC GGC AAC CTG CTG CTG CAG TAC GGC AGC TTC
 TGC ACC CAG CTG AAC CGG GGC CTG AGC GGC ATC GGC
 GGC CAG GAC GAC CGG AAC ACC CGG GAG GTG TTC GGC
 CAG GTG AAG CAG ATG TAC AAG ACC CCC ACC CTG AAG
 TAC TTC GGC GGC TTC AAC TTC AGC CAG ATC CTG CCC
 GAC CCC CTG AAG CCC ACC AAG GGC AGC TTC ATC GAG
 GAC CTG CTG TTC AAC AAG GTG ACC CTG GGC GAC GGC
 GGC TTC ATG AAG CAG TAC GGC GAG TGC CTG GGC GAC
 ATC AAC GGC CGG GAC CTG ATC TGC GGC CAG AAG TTC
 AAC GGC CTG ACC GTG CTG CCC CCC CTG CTG ACC GAC
 GAC ATG ATC GGC GGC TAC ACC GGC CCC CTG GTG AGC
 GGC ACC GGC ACC GGC GGC TGG ACC TTC GGC GGC GGC
 GGC GGC CTG CAG ATC CCC TTC GGC ATG CAG ATG GGC
 TAC CGG TTC AAC GGC ATC GGC GTG ACC CAG AAC GTG
 CTG TAC GAG AAC CAG AAG CAG ATC GGC AAC CAG TTC
 AAC AAG GGC ATC AGC CAG ATC CAG GAG AGC CTG ACC
 ACC ACC AGC ACC GGC CTG GGC AAG CTG CAG GAC GTG
 GTG AAC CAG AAC GGC CAG GGC CTG AAC ACC CTG GTG
 AAG CAG CTG AGC AGC AAC TTC GGC GGC ATC AGC AGC
 GTG CTG AAC GAC ATC CTG AGC CGG CTG GAC AAG GTG
 GAG GGC GAG GTG CAG ATC GAC CGG CTG ATC ACC GGC
 CGG CTG CAG AGC CTG CAG ACC TAC GTG ACC CAG CAG
 CTG ATC CGG GGC GGC GAG ATC CGG GGC ACC GGC AAC
 CTG GGC GGC ACC AAG ATG AGC GAG TGC GTG CTG GGC
 CAG AGC AAG CGG GTG GAC TTC TGC GGC AAG GGC TAC

-continued

CAC CTG ATG AGC TTC CCC CAG GCC GCC CAC GGC
 GTG GTG TTC CTG CAC GTG ACC TAC GTG CCG AGC CAG
 GAG CGG AAC TTC ACC ACC GCC CCC GCC ATC TGC CAC
 GAG GGC AAG GCC TAC TTC CCC CGG GAG GGC GTG TTC
 GTG TTC AAC GGC ACC CAG TGG TTC ATC ACC CAG CGG
 AAC TTC TTC AGC CCC CAG ATC ATC ACC ACC GAC AAC
 ACC TTC GTG AGC GGC AAC TGC GAC GTG GTG ATC GGC
 ATC ATC AAC AAC ACC GTG TAC GAC CCC CTG CAG CCC
 GAG CTG GAC AGC TTC AAG GAG GAG CTG GAC AAG TAC
 TTC AAG AAC CAC ACC ACC CCC CAG GTG GAC CTG GGC
 GAC ATC AGC GGC ATC AAC GCC AGC GTG GTG AAC ATC
 CAG AAG GAG ATC GAC CGG CTG AAC GAG GTG GCC AAG
 AAC CTG AAC GAG AGC CTG ATC GAC CTG CAG GAG CTG
 GGC AAG TAC GAG CAG TAC ATC AAG TGG CCC TGG

[0174] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:8 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:8 as follows: about 36 of the 79 phenylalanine codons are TTT, and about 43 of the phenylalanine codons are TTC; about 7 of the 92 leucine codons are TTA, about 12 of the leucine codons are TTG, about 12 of the leucine codons are CTT, about 18 of the leucine codons are CTC, about 7 of the leucine codons are CTA, and about 36 of the leucine codons are CTG; about 26 of the 73 isoleucine codons are ATT, about 35 of the isoleucine codons are ATC, and about 12 of the isoleucine codons are ATA; the 19 methionine codons are ATG; about 16 of the 89 valine codons are GTT, about 41 of the valine codons are GTG, about 11 of the valine codons are GTA, and about 21 of the valine codons are GTC; about 17 of the 93 serine codons are TCT, about 20 of the serine codons are TCC, about 14 of the serine codons are TCA, about 5 of the serine codons are TCG, about 15 of the serine codons are AGT, and about 22 of the serine codons are AGC; about 16 of the 57 proline codons are CCT, about 19 of the proline codons are CCC, about 16 of the proline codons are CCA, and about 6 of the proline codons are CCG; about 23 of the 94 threonine codons are ACT, about 34 of the threonine codons are ACC, about 26 of the threonine codons are ACA, and about 11 of the threonine codons are ACG; about 22 of the 84 alanine codons are GCT, about 34 of the alanine codons are GCC, about 19 of the alanine codons are GCA, and about 9 of the alanine codons are GCG; about 23 of the 52 tyrosine codons are TAT and about 29 of the tyrosine codons are TAC; about 6 of the 14 histidine codons are CAT and about 8 of the histidine codons are CAC; about 14 of the 55 glutamine codons are CAA and about 41 of the glutamine codons are CAG; about 37 of the 81 asparagine codons are AAT and about 44 of the asparagine codons are AAC; about 24 of the 57 lysine codons are AAA and about 33 of the

lysine codons are AAG; about 33 of the 71 aspartic acid codons are GAT and about 38 of the aspartic acid codons are GAC; about 17 of the 40 glutamic acid codons are GAA and about 23 of the glutamic acid codons are GAG; about 15 of the 33 cysteine codons are TGT and about 18 of the cysteine codons are TGC; the 10 tryptophan codons are TGG; about 3 of the 41 arginine codons are CGT, about 8 of the arginine codons are CGC, about 5 of the arginine codons are CGA, about 8 of the arginine codons are CGG, about 9 of the arginine codons are AGA, and about 8 of the arginine codons are AGG; and about 13 of the 77 glycine codons are GGT, about 26 of the glycine codons are GGC, about 19 of the glycine codons are GGA, and about 19 of the glycine codons are GGG.

[0175] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0176] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:8, optimized according to codon usage in humans is presented herein as SEQ ID NO:30.

ATG GAT GCA ATG AAG CGG GGC CTG TGC TGC GTG CTC
 CTG TGC TGC GGG GGG GTG TTT GTG AGC CCC AGT GCC
 AGA GGT AGC GGC AGC GAT TTG GAT AGG TGC ACC ACA
 TTT GAT GAC GTG CAG GCT CCC AAT TAC ACC CAG CAC
 ACC AGT TCT ATG AGA GGA GTA TAC TAC CCT GAG GAG
 ATC TTC CGC AGT GAT ACC CTA TAT TTA ACA CAA GAT
 TTA TTC TTA CCC TTC TAC TCC AAC GTC ACA GGG TTT
 CAC ACC ATT AAC CAC ACC TTC GGC AAC CCC GTG ATC
 CGG TTT AAA GAT GGC ATT TAT TTC GCA GCC ACA GAT
 AAG TCG AAT GTA GTG CGG GGT TGG GTT TTT GGA TCA
 ACA ATG AAT AAT AAA TCT CAG TCC GTG ATC ATT ATT
 AAC AAC TCT ACG AAT GTG GGT ATA GGA GCC TGT AAT
 TTC GAG TTA TGC GAT AAT CCA TTT TTC GGC GTG AGT
 AAA CCA AAG GGC ACT CAG ACC CAT AGC ATG ATT TTC
 GAT AAC GCA TTC AAT TGT AGC TTT GAA TAC ATT TCT
 GAT GCT TTT TCA CTC GAC GTT TCA GAA AAG TCT GGG
 AAC TTC AAG CAT TTA AGA GAG TGC GTC TTT AAA AAT
 AAC GAG GGG TTC CTG TAC GTG TAT AAA GGA TAC CAG
 CCT ATC GAC GTG GTG CGG GAG CAG CTC GCA AGC GGT TTT
 AAT ACC CTG AAG CCC ATC TTT AAG GTG CCC CTG GGA
 ATC AAT ATT ACA AAC TTC AGG GCT ATC CTC ACC GCT
 TTT AGC CCA GCT CAG GAC ATA TGG GAC ACC TCC GCC

-continued

GCC GGC TAC TTC CTC GGA TAT TTG AAA CCA ACC ACA
 TTC ATG CTG AAG TAT GAC GAA AAT GGG ACG ATT ACC
 GAC GGC GTA GAC TGT AGT CAG AAC CCT TTG GCG GAG
 TTG AAG TGC TCA CTC AAG AGC TTT GAG ATC GAC AAG
 GGA ATT TAT CAA ACT AGC AAC TTC AGG GTG GTG CCC
 TCC GGA GAT GTA GTT CGC TTC CCC AAC ATC ACC AAC
 CTG TGC CCG TTC GGT GAG GTG TTT AAT GCA ACT AAA
 TTC CCC TCA GTG TAT GCG TGG GAA AGA AAG AAA ATT
 AGC AAC TGT GTT GCC GAT TAC AGC GTC CTT TAT AAC
 TCA ACA TTC TTC TCT ACC TTT AAG TGC TAT GGT GTG
 TCC GGC ACT AAG TTG AAC GAC CTC TGC TTT AGT AAC
 GTG TAC GCT GAT TCC TTG GTG GTG AAA GGG GAT GAC
 GTG GGT CAG ATT GCA CGC GGC CAG ACC GGA GTA ATC
 GCC GAT TAC AAT TAC AAG TTG COT GAC GAC TTC ATG
 GGC TGC GTT CTA GCA TGG AAT ACC CGC AAC ATA GAT
 GCC ACC TCA ACG GGG AAC TAC AAC TAC AAG TAC AGA
 TAT CTG AGA CAC GGT AGC CTG CGG CCT TTT GAG CGG
 GAT ATC TCC AAT GTG COT TTT AGC CCC GAT GGC AAA
 CCA TGC ACC CCA COT GGC CTG AAT TGT TAT TGG CCT
 TTG AAC GAT TAT GGA TTC TAC ACT ACC ACT GGG ATC
 GGT TAT CAA CCC TAC CGG GTC GTC GTC CTG AGT TTT
 GAA CTC TTG AAC GCG CCT GCA ACA GTC TGC GGA CCC
 AAG CTG TCG ACA GAC CTT ATC AAG AAT CAG TGT GTG
 AAC TTT AAC TTC AAT GGG CTC ACC GGT ACC GGT GTT
 CTG ACT CCA TCT AGT AAG CGA TTT CAA CCA TTC CAA
 CAG TTC GGC GGT GAC GTT TCC GAT TTT ACG GAT TCG
 GTG GGT GAT CCA AAA ACA TCA GAG ATC CTT GAC ATA
 TCG CGG TGT TCT TTT GGA GGC GTG TCT GTG ATT ACA
 CCA GGC ACT AAT GCT AGT AGC GAA GTC GCT GTA CTA
 TAC CAG GAC GTG AAC TGC ACC GAC GTG AGC ACG GCA
 ATC CAC GCT GAC CAG CTG ACC CCC GGC TGG GGC ATC
 TAC AGT ACA GGC AAT AAC GTC TTT CAG ACC CAG GCC
 GGC TGT CTG ATT GGG GCT GAG CAG GTC GAC ACT TCC
 TAT GAA TGT GAT ATT CCC ATC GGC GCT GGA ATT TGT
 GCT AGC TAT CAC ACA GTC TCC CTT TTA AGA TCA ACC
 AGC CAG AAA TCT ATT GTG GCT TAC ACA ATG TCT CTC
 GGC GCA GAC TCA TCA ATT GCC TAT AGC AAC AAT ACC
 ATT GCA ATC CCT ACC AAT TTT AGT ATA TCC ATA ACC

-continued

ACC GAG GTG ATG CCC GTG TCT ATG GCG AAA ACT TCC
 GTC GAT TCC AAC ATG TAT ATC TCG GGG GAC TCC ACA
 GAA TGC GCC AAC CTG CTT CTG CAG TAT GGA AGC TTC
 TGT ACT CAA CTC AAC CGC GCA TTG TCT GGG ATT GCC
 GCC GAG CAG GAT AGG AAT ACT AGA GAG GTG TTC GGT
 CAG GTT AAA CAA ATG TAC AAG ACA CCG ACA CTT AAG
 TAC TTC GGA GGT TTT AAC TTT TCC CAG ATA CTC CCT
 GAC CCT CTA AAG CCT ACT AAA CGC AGT TTC ATC GAG
 GAT CTC CTG TTT AAT AAG GTG ACA CTC GCC GAT GCT
 GGC TTC ATG AAA CAA TAC GGA GAA TGC GTG GGA GAC
 ATT AAC GCC AGA GAC CTG ATC TCT GCC CAG AAG TTC
 AAC GGT CTG ACA GTA CTT CCT CCC CTT CTG ACG GAC
 GAC ATG ATT GCT GCA TAC ACA GCC CCA CTA GTT AGC
 GGC ACA GCC ACA GCT GGG TGG ACC TTT GCG GCT GGC
 GCA GCG TTG CAG ATT CCA TTC GCG ATG CAG ATG GCT
 TAC CCA TTT AAC GGG ATC GGC GTG ACT CAG AAT GTT
 TTG TAT GAG AAC CAG AAA CAG ATC GCT AAT CAG TTT
 AAC AAG GCA ATC AGC CAG ATA CAA GAA TCT CTG ACT
 ACC ACA AGC ACC GCT CTG GGA AAA CTG CAG GAC GTG
 GTG AAT CAG AAT GCA CAG GCC CTC AAC ACG CTC GTG
 AAG CAG CTT AGT TCC AAT TTC GGG GCG ATC TCC TCC
 GTT TTA AAT GAT ATC CTG AGT GCT CTG GAC AAG GTC
 GAG GCC GAA GTT CAG ATC GAC CGC CTG AAT ACA GGG
 AGG CTA CAA TCA TTG CAG ACT TAC GTG ACT CAG CAG
 CTC ATA AGG GCT GCA GAG ATT AGG GCC TCT GCA AAC
 CTT GCG GCG ACC AAG ATG TCC GAG TGT GTT CTC GGT
 CAG TCC AAA CGG GTT GAC TTT TGT GGC AAA GGC TAC
 CAT CTG ATG AGC TTC CCC CAG GCC GCA CCC CAT GGC
 GTA CTT TTT CTG CAC GTA ACT TAT GTG CCA TCC CAA
 GAA AGG AAC TTC ACT ACG GCG GCA GCC ATA TGC CAT
 GAA GGT AAA GCA TAT TTC COT CCA GAA GGG GTA TTT
 GTT TTC AAC GGG ACT AGC TGG TTT ATT ACG CAG CGG
 AAT TTC TTT TCA CCA CAA ATC ATC ACT ACT GAT AAC
 ACA TTC GTC AGC GGC AAT TGT GAT GTC GTC ATT GGA
 ATT ATA AAC AAC ACT GTG TAC GAT COT CTG CAG CGG
 GAA CTG GAT TCT TTT AAG GAG GAG GTC GAC AAG TAC
 TTC AAA AAC CAT ACC TCG CCC GAC GTG GAC CTA GGC
 GAT ATC TCT GGG ATT AAT GCC TCA GTA GTC AAC ATC
 CAG AAG GAG ATA GAC CGA CTT AAT GAG GTT GCC AAG

-continued

AAT CTG AAT GAG AGT CTC ATC GAT CTG CAA GAA CTT
GGC AAG TAT GAA CAA TAT ATC AAA TGG CCA TGG

[0177] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:10 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO: 10 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:10, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:10 is shown in Table 13.

TABLE 13

AMINO ACID	Number in SEQ ID NO: 10
A	Ala 41
R	Arg 25
C	Cys 23
G	Gly 47
H	His 9
I	Ile 37
L	Leu 46
K	Lys 32
M	Met 9
F	Phe 51
P	Phe 38
S	Ser 58
T	Thr 56
W	Trp 6
Y	Tyr 35
V	Val 56
N	Asn 46
D	Asp 45
Q	Gln 21
E	Glu 17

[0178] Using the amino acid composition shown in Table 13, a human codon-optimized coding region which encodes SEQ ID NO:10 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:10 as follows: the 51 phenylalanine codons are TTC, the 46 leucine codons are CTG, the 37 isoleucine codons are ATC, the 9 methionine codons are ATG, the 56 valine codons are GTG, the 58 serine codons are AGC, the 38 proline codons are CCC, the 56 threonine codons are ACC, the 41 alanine codons are GCC, the 35 tyrosine codons are TAC, the 9 histidine codons are CAC, the 21 glutamine codons are CAG, the 46 asparagine codons are AAC, the 32 lysine codons are AAG, the 45 aspartic acid codons are GAC, the 17 glutamic acid codons are GAG, the 23 cysteine codons are TGC, the 6 tryptophan codons are TGG, the 25 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 47 glycine codons are GGC. The codon-optimized coding region designed by this method is presented herein as SEQ ID NO:33.

ATG GAC GCC ATG AAG CGG GGC CTG TGC TGC GTG CTG
CTG CTG TGC GGC GGC GTG TTC CTG AGC CCC AGC GGC
CGG GGC AGC GGC AGC GAC CTG GAC CGG TGC ACC ACC
TTC GAC GAC GTG CAG GGC CCC AAC TAC ACC CAG CAC
ACC AGC AGC ATG CGG GGC GTG TAC TAC CCC GGC AGC GAG
ATC TTC CGG AGC GAC ACC CTG TAC CTG ACC CAG GAC
CTG TTC CTG CCC TTC TAC AGC AAC GTG ACC GGC TTC
CAC ACC ATC AAC CAC ACC TTC GGC AAC CCC GTG ATC
CCC TTC AAG GAC GGC ATC TAC TTC GGC GGC ACC GAG
AAG AGC AAC CTG GTG CTG CGG GGC TGG GTG TTC GGC AGC
ACC ATG AAC AAC AAG AGC CAG AGC GTG ATC ATC ATC
AAC AAC AGC ACC AAC GTG GTG ATC CGG GGC TGC AAC
TTC GAG CTG TGC GAC AAC CCC TTC TTC GGC GTG AGC
AAG CCC ATG GGC ACC CAG ACC CAC ACC ATG ATC TTC
GAC AAC GCC TTC AAC TGC ACC TTC GAG TAC ATC AGC
GAC GCC TTC AGC CTG GAC GTG AGC GAG AAG AGC GGC
AAC TTC AAC CAG CTG CGG GAG TTC GTG TTC AAG AAC
AAG GAC GGC TTC CTG TAC GTG TAC AAG GGC TAC CAG
CCC ATC GAC GTG GTG CGG GAC CTG CCC AGC GGC TTC
AAC ACC CTG AAG CCC ATC TTC AAG CTG CCC CTG GGC
ATC AAC ATC ACC AAC TTC CGG GGC ATC CTG ACC GGC
TTC AGC CCC GGC CAG GAC ATC TGG GGC ACC AGC GGC
GCC GCC TAC TTC GTG GGC TAC CTG AAG CCC ACC ACC
TTC ATG CTG AAG TAC GAC GAG AAC GGC ACC ATC ACC
GAC GCC CTG GAC TGC AGC CAG AAC CCC CTG GGC GAG
CTG AAG TGC AGC GTG AAG AGC TTC GAG ATC GAC AAG
GGC ATC TAC CAG ACC AGC AAC TTC CCG GTG GTG CCC
AGC GGC GAC GTG GTG CGG TTC CCC AAC ATC ACC AAC
CTG TGC CCC TTC GGC GAG GTG TTC AAC GGC ACC AAG
TTC CCC AGC GTG TAC GGC TGG GAG CGG AAG AAG ATC
AGC AAC TGC GTG GGC GAC TAC AGC GTG CTG TAC AAC
AGC ACC TTC TTC AGC ACC TAC AAG TGC TAC GGC GTG
AGC GGC ACC AAG CTG AAC GAC CTG TGC TTC AGC AAC
GTG TAC GCC GAC AGC TTC GTG GTG AAG GGC GAC GAC
GTG CGG GAC ATC GGC CCC GGC CAG ACC GGC GTG ATC
GCC GAC TAC AAC TAC AAG CTG GGC ACC GAC TTC ATG
GGC TGC CTG GTG GGC TGG AAC ACC CGG AAG ATC GAC
GCC ACC AGC ACC GGC AAC TAC AAC TAC AAG TAC CGG
TAC CTG CGG CAC GGC AAG CTG CGG CCC TTC GAG CGG

-continued-

GAC ATC AGC AAC GTG CCC TTC AGC CCC GAC GGC AAG
 CCC TGC ACC CCC CCC GGC CTG AAC TGC TAC TGC CCC
 CTG AAC GAC TAC GGC TTC TAC ACC ACC ACC GGC ATC
 GGC TAC CAG CCC TAC CGG GTG GTG GTG CTG AGC TTC
 AAG CTG AGC ACC GAC CTG ATC AAG AAC CAG TGC GTG
 AAC TTC AAC TTC AAC GGC CTG ACC GGC ACC GGC GTG
 CTG ACC CCC AGC AGC AAG GGC TTC CAG CCC TTC CAG
 CAG TTC GGC CGG GAC CTG AGC GAC TTC ACC GAC AGC
 CTG CGG GAC CCC AAG AAC AGC GAG ATC CTG GAC ATC
 AGC CCC TGC AGC TTC GGC GGC GTG AGC GTG ATC ACC
 CCC GGC ACC AAC GGC AGC AGC GAG GTG GCC GTG CTG
 TAC CAG GAC GTG AAC TGC ACC GAC GTG AGC ACC GGC
 ATC CAC GGC GAC CAG CTG ACC CCC GGC TGG CGG ATC
 TAC AGC ACC GGC AAC AAC CTG TTC CAG ACC CAG GGC
 GGC TGC CTG ATC GGC GGC GAC CAG GTG GAC ACC AGC
 TAC GAG TGC GAC ATC CCC ATC GGC GGC GGC ATC TGC
 GCC AGC TAC CAC ACC CTG AGC CTG CTG CGG AGC ACC
 GGC CAG AAG AGC ATC GTG GCC TAC ACC ATG AGC CTG
 GGC

[0179] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:10 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:10 as follows: about 23 of the 51 phenylalanine codons are TTT, and about 28 of the phenylalanine codons are TTC; about 3 of the 46 leucine codons are TTA, about 6 of the leucine codons are TTG, about 6 of the leucine codons are CTT, about 9 of the leucine codons are CTC, about 4 of the leucine codons are CTA, and about 18 of the leucine codons are CTG; about 13 of the 37 isoleucine codons are ATT, about 18 of the isoleucine codons are ATC, and about 6 of the isoleucine codons are ATA; the 9 methionine codons are ATG; about 10 of the 56 valine codons are GTT, about 26 of the valine codons are GTG, about 7 of the valine codons are GTA, and about 13 of the valine codons are GTC; about 11 of the 58 serine codons are TCT, about 13 of the serine codons are TCC, about 9 of the serine codons are TCA, about 3 of the serine codons are TCG, about 8 of the serine codons are AGT, and about 14 of the serine codons are AGC; about 11 of the 38 proline codons are CCT, about 13 of the proline codons are CCC, about 10 of the proline codons are CCA, and about 4 of the proline codons are CCG; about 14 of the 56 threonine codons are ACT, about 20 of the threonine codons are ACC, about 16 of the threonine codons are ACA, and about 6 of the threonine codons are ACG; about 11 of the 41 alanine

codons are GCT, about 16 of the alanine codons are GCC, about 10 of the alanine codons are GCA, and about 4 of the alanine codons are GCG; about 15 of the 35 tyrosine codons are TAT and about 20 of the tyrosine codons are TAC; about 4 of the 9 histidine codons are CAT and about 5 of the histidine codons are CAC; about 5 of the 21 glutamine codons are CAA and about 16 of the glutamine codons are CAG; about 21 of the 46 asparagine codons are AAT and about 25 of the asparagine codons are AAC; about 14 of the 32 lysine codons are AAA and about 18 of the lysine codons are AAG; about 21 of the 45 aspartic acid codons are GAT and about 24 of the aspartic acid codons are GAC; about 7 of the 17 glutamic acid codons are GAA and about 10 of the glutamic acid codons are GAG; about 10 of the 23 cysteine codons are TGT and about 13 of the cysteine codons are TGC; the 6 tryptophan codons are TGG; about 2 of the 25 arginine codons are CGT, about 5 of the arginine codons are CGC, about 3 of the arginine codons are CGA, about 5 of the arginine codons are CGG, about 5 of the arginine codons are AGA, and about 5 of the arginine codons are AGG; and about 8 of the 47 glycine codons are GGT, about 16 of the glycine codons are GGC, about 11 of the glycine codons are GGA, and about 12 of the glycine codons are GGG.

[0180] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a given amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0181] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO: 10, optimized according to codon usage in humans is presented herein as SEQ ID NO:32.

ATG GAC GCC ATG AAG CGA GGA CTG TGC TGT GTT TTG
 TCG TGC TGC GGC GCA GTT TTT GTC AGT CCA TCC GCC
 CGG GGG TGC GGA TCT GAC CTA AGT GAA TGC ACG ACC
 TTC GAT GAC GTG CAG GCA CCA AAT TAC ACC CAA CAT
 ACT TCA TCC NAT GCG GGC GTT TAC TAC CCC GAC GAA
 ATC TTC GGG AGT GAC ACC CTG TAT GTC ACT CAG GAC
 CTG TTT CTG CCC TTC TAC AGC AAT GTG ACA GGC TTT
 CAC ACC AAT AAC CAT ACC TTC GGG AAT CCA GTA ATC
 CCT TTT AAG GAT GGG ATT TAC TTT GCT GCT ACT GAG
 AAA AGT AAT GTT GTC AGG GGG TGG TTT TTT GGC TCA
 ACA ATG AAC AAT AAG TCT CAG AGT GTC ATC ATT
 AAC AAT TCT AAT AAC GTA GTC ATC AGA GCA TGC AAC
 TTC GAG CTC TGT GAT AAC CCT TTC TTT GCT GTG TCT
 AAG CCC ATG GGC ACT CAA ACA CAT ACC ATG ATC TTC
 GAC AAT GCG TTC AAT TGT ACC TTT GAG TAT ATA TCA
 GAC GCC TTC AAG CTA GAC CTC TCG GAA AAG TCC GGA

-continued

AAC TTT AAA CAC CTC CGG GAA TTC GTG TTT AAG AAC
 AAA GAT GGA TTT TTG TAC GTA TAC AAG GGT TAT CAG
 CCT ATC GAT GTC GTG GGT GAT CTG CCC TCC GGC TTC
 AAC ACC CTG AAG CCT ATA TTC AAA CTA CCC CTA GGG
 ATC AAC ATC ACC AAT TTT AGG GCA ATA CTT ACG GCA
 TTT TCC CCA GCC CAG GAC ATC TGG GGA ACT TCC GCC
 GCT GCC TAC TTT GTG GGC TAT CTC AAG CTT ACT ACT
 TTC ATG CTT AAG TAT GAT GAG AAT GGC ACA ATC ACG
 GAT GCA GTG GAT TGC TCG CAG AAT CCA CTT GCT GAG
 CTG AAA TGC TCC GTA AAG AGC TTC GAA ATT GAT AAA
 GGA ATC TAT CAG ACC AGC AAC TTC CGG GTC GTG CCC
 TCT GGC GAC GTT GTC CGG TTC CCC AAC ATC ACC AAC
 CTC TGC CCA TTC GGC GAG GTG TTC AAC OCT ACA AAA
 TTC CCA AGT GTC TAC GCC TGG GAG AGG AAA AAG ATC
 TCT AAT TGT GTG GCA GAT TAT TCC GTG TTA TAC AAC
 AGC ACA TTC TTC TCA ACG TTC AAG TGT TAT GGC GTG
 AGC GCC ACC AAG CTT AAC GAC CTC TGC TTC TCC AAT
 GTA TAC GCT GAC TCT TTT GTG GTT AAG GGA GAC GAT
 GTG CGA CAG ATC GCC CGC GGG CAA ACC GGA GTG ATT
 GCG GAC TAC AAC TAT AAA CTG CCC GAC GAT TTC ATG
 GGT TGT GTG CTT GCT TGG AAT ACG AGG AAC ATT GAC
 GCA ACG AGC ACC GGG AAC TAT AAT TAC AAA TAT CGT
 TAC CTG CGC CAT GGG AAA CTC AGA OCT TTT GAA CGA
 GAT ATT AGC AAC GTC CTT TTC TCA CGG GAT GGG AAG
 CCC TGT ACC CCA OCT GCC CTC AAG TGC TAT TGG CCT
 CTC AAC GAC TAC GGC TTC TAC ACT AAC ACA GGG ATC
 GGC TAC CAG CCC TAT CGC GTG GTG GTT CTC TCC TTT
 GAA CTC CTT AAT GCT CCC GCG ACT GTG TGT GGG CCG
 AAG TTG AGT ACT GAC TTA ATA AAA AAT CAA TGC GTA
 AAC TTT AAC TTT AAT GGC TTC ACA GGT ACA GGT GTG
 CTC ACA CGG AGT AGC AAA AGG TTC CAG CCA TTT CAG
 CAA TTT GGC AGA GAT GTG TCT GAC TTT ACA GAC AGC
 GTG GGC GAT CTT AAG ACT TCT GAG ATT TTA GAC ATC
 TCA CCT TGT TCC TTT GGA GGA GTG AGC GTG ATA ACT
 CCC GGT ACC AAC GCC TCA TCC GAA GTG GCT GTC CTG
 TAT CAG GAC GTT AAT TGC AAC GAT GTC TCT ACA GGC
 ATT CAC GCC GAT CAG CTC ACA CCA GCT TGG CAC ATC
 TAC AGT ACC GGT AAC AAT GTT TTC CAG ACT CAG GCC
 GGT TGT CTG ATT GGC GCC GAG CAG GTC GAC ACA TCT

-continued

TAC GAG TGC GAT ATT CCC ATA GGT GCC GGC ATT TGT
 GCG AGC TAC CAC ACT GTA TCA CTG CTG AGA AGC ACA
 AGC CAG AAA TCA ATT GTG GCA TAC ACA ATG TCC TTG
 GGA GCA

[0182] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:12 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:12 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:12, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:12 is shown in Table 14.

TABLE 14

AMINO ACID	Number in SEQ ID NO: 12	
A	Ala	46
R	Arg	18
C	Cys	13
G	Gly	34
H	His	5
I	Ile	36
L	Leu	50
K	Lys	26
M	Met	12
F	Phe	29
P	Pro	20
S	Ser	38
T	Thr	38
W	Trp	4
Y	Tyr	17
V	Val	36
N	Asn	35
D	Asp	27
Q	Gln	34
E	Glu	23

[0183] Using the amino acid composition shown in Table 14, a human codon-optimized coding region which encodes SEQ ID NO:12 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:12 as follows: the 29 phenylalanine codons are TTC, the 50 leucine codons are CTG, the 36 isoleucine codons are ATC, the 12 methionine codons are ATG, the 36 valine codons are GTG, the 38 serine codons are AGC, the 20 proline codons are CCC, the 38 threonine codons are ACC, the 46 alanine codons are GCC, the 17 tyrosine codons are TAC, the 5 histidine codons are CAC, the 34 glutamine codons are CAG, the 35 asparagine codons are AAC, the 26 lysine codons are AAG, the 35 aspartic acid codons are GAC, the 23 glutamic acid codons are GAG, the 13 cysteine codons are TGC, the 4 tryptophan codons are TGG, the 18 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 34 glycine codons are GGC.

The codon-optimized coding region designed by this method is presented herein as SEQ ID NO:35.

ATG GAC GCC ATG AAG CGG GGC CTG TGC TGC GTG CTG
CTG CTG TGC GGC GCC CTG TTC GTG AGC CCC AGC GGC
CGG GGC AGC GGC GAC AGC AGC ATC GCC TAC AGC AAC
AAC ACC ATC GCC ATC CCC AAC AAC TTC AGC ATC AGC
ATC ACC ACC GAG GTG ATG CCC GTG AGC ATG GCC AAG
ACC AGC GTG GAC TGC AAC ATG TAC ATC TGC GGC GAC
AGC ACC GAG TGC GCC AAC CTG CTG CTG CAG TAC GGC
AGC TTC TGC ACC CAG CTG AAC CGG GCC CTG AGC GGC
ATC GCC GCC GAG CAG GAC CGG AAC ACC CGG GAG GTG
TTC GCC CAG GTG AAG CAG ATG TAC AAG ACC CCC ACC
CTG AAG TAC TTC GGC GGC TTC AAC TTC AGC CAG ATC
CTG CCC GAC CCC CTG AAG CCC ACC AAG CGG AGC TTC
ATC GAG GAC CTG CTG TTC AAC AAG GTG ACC CTG GCC
GAC GCC GGC TTC ATG AAG CAG CAG TAC GGC GAG TGC CTG
GGC GAC ATC AAC GCC CGG GAC CTG ATC TGC GCC CAG
AAG TTC AAC GGC CTG ACC GTG CTG CCC CCC CTG CTG
ACC GAC GAC ATG ATC GCC GCC TAC ACC GCC GCC CTG
GTG AGC GGC ACC GCC ACC CGC GGC TGC ACC TTC GGC
GCC GGC GCC GCC CTG CAG ATC CCC TTC GCC ATG CAG
ATG GCC TAC CGG TTC AAC GGC ATC GGC GTG ACC CAG
AAC GTG CTG TAC GAG AAC CAG AAG CAG ATC GCC AAC
CAG TTC AAC AAG GCC ATC AGC CAG ATC CAG GAG AGC
CTG ACC ACC ACC AGC ACC GCC CTG GGC AAG CTG CAG
GAC GTG GTG AAC CAG AAC GCC CAG GCC CTG AAC ACC
CTG GTG AAG CAG CTG AGC AGC AAC TTC GGC GCC ATC
AGC AGC GTG CTG AAC GAC ATC CTG AGC CGG CTG GAC
AAG GTG GAG GGC AAG GTG CAG ATC GAC GGC CTG ATC
ACC GGC CGG CTG CAG AGC CTG CAG ACC TAC GTG ACC
CAG CAG CTG ATC CGG GGC GGC GAG ATC CGG GCC AGC
GCC AAC CTG GCC ACC AAC AAG ATG AGC GAG TGC GTG
CTG GGC CAG AAG AAG CGG GTG CAG ATC TTC GGC AGC
GGC TAC CAC CTG ATG AAG TTC CCC CAG GCC GCC CCC
CAC GGC GTG GTG TTC CTG CAC GTG ACC TAC GTG CCC
AGC CAG GAG CGG AAC TTC ACC ACC GCC CCC GCC ATC
TGC CAG GAG GGC AAG GCC TAC ACC ACC CGG GAG GGC
GTG TTC GTG TTC CAG GGC GGC ACC TCG TTC ATC ACC
CAG CGG AAC TTC TTC ACC GCC CAG ATC ATC ACC ACC

-continued

GAC AAC ACC TTC GTG AGC GGC AAC TGC GAG GTG GTG
ATC GGC ATC ATC AAC AAC ACC CTG TAC GAC CCC CTG
CAG CCC GAG CTG GAC AGC TTC AAG GAG GAG CTG GAC
AAG TAC TTC AAG AAC CAC ACC ACC CCC GAG GTG GAC
CTG GGC GAC ATC AGC GGC ATC AAC GCC AGC GTG GTG
AAC ATC CAG AAG GAG ATC GAC CGG CTG AAC GAG GTG
GCC AAG AAC CTG AAC GAG AGC CTG ATC GAC CTG CAG
GAG CTG GGC AAG TAC GAG CAG TAC ATC AAG TGG CCC
TGG

[0184] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:12 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:12 as follows: about 13 of the 29 phenylalanine codons are TTT, and about 16 of the phenylalanine codons are TTC; about 4 of the 50 leucine codons are TTA, about 6 of the leucine codons are TTG, about 6 of the leucine codons are CTT, about 10 of the leucine codons are CTC, about 4 of the leucine codons are CTA, and about 20 of the leucine codons are CTG; about 13 of the 36 isoleucine codons are ATT, about 17 of the isoleucine codons are ATC, and about 6 of the isoleucine codons are ATA; the 12 methionine codons are ATG; about 6 of the 36 valine codons are GTT, about 9 of the valine codons are GTG, about 4 of the valine codons are GTA, and about 17 of the valine codons are GTG; about 7 of the 38 serine codons are TCT, about 8 of the serine codons are TCC, about 6 of the serine codons are TCA, about 2 of the serine codons are TCG, about 6 of the serine codons are AGT, and about 9 of the serine codons are AGC; about 6 of the 20 proline codons are CCT, about 7 of the proline codons are CCC, about 5 of the proline codons are CCA, and about 2 of the proline codons are CCG; about 9 of the 38 threonine codons are ACT, about 14 of the threonine codons are ACC, about 11 of the threonine codons are ACA, and about 4 of the threonine codons are ACG; about 12 of the 46 alanine codons are GCT, about 19 of the alanine codons are GCC, about 10 of the alanine codons are GCA, and about 5 of the alanine codons are GCG; about 7 of the 17 tyrosine codons are TAT and about 10 of the tyrosine codons are TAC; about 2 of the 5 histidine codons are CAT and about 3 of the histidine codons are CAC; about 9 of the 34 glutamine codons are CAA and about 25 of the glutamine codons are CAG; about 16 of the 35 asparagine codons are AAT and about 19 of the asparagine codons are AAC; about 11 of the 26 lysine codons are AAA and about 15 of the lysine codons are AAG; about 12 of the 27 aspartic acid codons are GAT and about 15 of the aspartic acid codons are GAC; about 16 of the 23 glutamic acid codons are GAA and about 13 of the glutamic acid codons are GAG; about 6 of the 13 cysteine codons are TGT and about 7 of the cysteine codons are TGC; the 4 tryptophan codons are TGG; about 1 of the 18 arginine codons are CGT, about 3 of the arginine codons are CGC, about 2 of the arginine codons are CGA, about 4 of the arginine codons are CGG, about 4 of the arginine codons are

AGA, and about 4 of the arginine codons are AGG; and about 6 of the 34 glycine codons are GGT, about 12 of the glycine codons are GGC, about 8 of the glycine codons are GGA, and about 8 of the glycine codons are GGG.

[0185] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0186] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:12, optimized according to codon usage in humans is presented herein as SEQ ID NO:34.

ATG GAT GCA ATG AAA AGA GGC CTG TGT TGT GTT CTG
CTG CTG TGT GGG GGG GTA TTT GTG AGT CCC TCT GCC
ACG GGA AGC GGC GAC AGC AGT ATA GCC TAC TCA AAC
AAT ACC ATC GCC ATT CCT ACA AAT TTT TCC ATC TCA
ATC ACG ACG GAA CTC ATG CCA GTT AGC ATG GCC AAA
ACC TCT GTC GAC TGC AAC ATG TAC ATC TGC GGA GAC
TCT ACT GAG TGC GCA AAC CTG CTC TTG CAG TAT GGC
TCG TTT TGC ACC CAG TAT AAT GGG GCC CTC AGT GGC
ATT GCC GCA GAA CAA GAT CGG AAT ACC AGG GAG GTC
TTC GCG CAA GTC AAG CAG ATG TAC AAA ACC CCT ACA
CTC AAA TAC TTC GGG GGG TTC AAC TTT AGC CAA ATC
CTG CCA GAC CCC CTC AAG CCT ACT AAG GGC AGT TTT
ATC GAA GAC TTA CTC TTT AAT AAG GTG ACA TTA GCT
GAT GCC GGA TTC ATG AAG CAG TAC GGA GAG TGC CTG
GGG GAT ATC AAC GCG CGG GAC CTA ATC TGT GCC CAG
AAG TTC AAC GGT CTC ACA GTG CTT CCG CCT CTC CTG
ACC GAT GAT ATG ATC GCA GCT TAC ACC GCC GCA CTG
GTT AGT GGT ACG GCC ACA GCA GGC TGG ACC TTC GGT
GCC GGT GCT GCC CTG CAA ATC CCA TTC GCG ATG CAG
ATG GCA TAC GAA TTT AAC GGC AAT GGA TTC ACC CAG
AAT GTC CTA TAC CAG AAC CAG AAG CAA ATC GCT AAC
CAG TTC AAC AAA GCC ATA TCC CAG ATT CAG GAG TCC
CTT ACT ACA ACC AGT ACT GCT TTA GGT AAA CTG CAA
GAT GTA GTG AAC CAG AAC GCT CAG GCC TTA AAT ACC
CTT GTT AAA CAG CTA TCC TCA GAC TTT GGG GCT ATC
TCC TCC GTG CTC AAC GAT ATC CTG AGC GGC CTC GAT
AAG GTG GAA GCG GAG CTC CAG ATC GAT AGA CTT ATT
ACA GGC AGG CTT CAG TCT CTC CAG ACC TAT GTC ACA

-continued

CAA CAG CTC ATT GGT GCT GCA GAG ATC GGC GCT TCC
GCC AAC TTG GCT GCA ACA AAG ATG TCT GAA TGT GTG
CTG GGA CAG AGC AAG AGA GTG CAG TTT TGT GGG AAA
GGC TAT CAC TTG ATG AGC TTC CCC CAG GCC GCC CCC
CAT GGA GTG GTA TTC CTA CAC GTG ACG TAC GTT CCA
TCT CAA GAA GGA AAT TTC ACC ACC GCA CCT GCC ATT
TGC CAC GAA GGG AAG GCT TAT TTC CCT CGA GAG GGC
GTG TTC GTT TTT AAC GGG ACT TCA TGG TTT ATA ACT
CAA AGG AAT TTC TTC TCG CCC CAG ATA ATT ACA ACA
GAC AAC ACT TTT GTG AGC GGC AAT TGC GAC GTG GTC
ATA GGT ATT ATT AAT AAT ACT GTG TAT GAC CCG CTG
CAG CCC GAA CTC GAC AGC TTT AAA GAG GAG CTG GAC
AAA TAC TTC AAG AAT CAT ACT TCA CCC CAG GTG GAT
CTG GGC CAG ATA TCC GGA ATC AAT GCC TCT GTG GTA
AAC ATT CAG AAG GAG ATC GAT GCG CTG AAG GAA GTG
GCT AAG AAT CTG AAT GAA TCA TTG ATT GAC CTT CAG
GAG TTG GGC AAG TAT GAG CAG TAT ATT AAA TGG CCA
TGG

[0187] Another representative codon-optimized coding region encoding SEQ ID NO:12 is presented herein as SEQ ID NO:47.

ATG GAT GCG ATG AAG GGA GGC CTG TGT TGC GTA CTG
CTG CTG TGC GGC GCC GTG TTT GTG AGC CCC AGC GCC
CGG GGC AGT GGC GAC AGC AGC AAT GCC TAT TGC AAC
GAC ACT AAT GCC ATA CCC ACA AAC TTC TCT ATA TCT
ATA ACT ACG GAG GTG ATG CCC GTT ATG GGC AAG
ACT AGT GTA GAC TGC AAC ATG TAC ATC TGC GGC GAC
TCT ACT GAG TGC GCC AAC CTG CTG CTG CAG TAT GGC
TCT TTC TGC ACC CAG CTG AAC AGA GCG CTG AGT GGC
ATC GCC GCG GAG CAG GAC CGG AAC ACA AGA GAG GTT
TTC GCC CAG GTA AAG CAG ATG TAC AAG ACC CCC ACT
CTG AAG TAT TTT GGC GGC TTC AAC TCT TCT CAG ATC
CTG CCC GAT CCC CTG AAG CCC ACC AAG AGG TCT TTC
ATC GAG GAG CTG CTG TTC AAC AAG GTC ACT CTG GCC
GAT GCC GGC TTC ATG AAG CAG TAC GGC GAG TGC CTG
GGC GAC ATT AAC GCC GGC GAC CTG ATC TGT GCC CAG
AAG TTT AAC GGC CTG AGC GTC CCC CCC CTG CTG
ACA GAT GAT ATG ATC GCC GCC TAC ACT GCC GCC CTG

-continued

GTC TCT GGC ACC GGC ACC GGC GGC TGG ACT TTC GGC
 GCC GGC GCC GGC CTC CAG ATC CCC TTC GCC ATG CAG
 ATG GCC TAT AGA TTT AAC GGC ATA GGC GTA ACT CAG
 AAC GTC CTG TAC GAG AAG CAG AAG CAG ATC GGC AAC
 CAG TTT AAC AAG GGC ATC TCC CAG ATT CAG GAG AGC
 CTG ACA ACC ACT AGC ACT GGC CTC GGC AAG CTG CAG
 GAC GTG GTG AAC CAG AAC GGC CAG GCC CTG AAC ACA
 CTG GTT AAG CAG CTG AGT TCT AAC TTT GGC GCC ATA
 TCC TCG GTG CTG AAC GAC ATA CTG TCA AGG CTG GAC
 AAG GTC GAG GCC GAG GTT CAG ATA GAT AGA CTG ATC
 ACA GGC AGA CTG CAG AGC CTG CAG ACC TAC GTT ACA
 CAG CAG CTG ATC AGA GGC GCC GAG ATC AGA GCC TCA
 GCC AAC CTG GCC GGC AAG ATG TCT GAG TGC GTC
 CTG GGC CAG TCT AAG AGA GTC GAT TTC TGC GGC AAG
 GGC TAC CAG CTG ATG AGT TTC CCC CAG GCC GGC CCC
 CAT GGC GTT GTA TTC CTG CAT GTG ACA TAT GTT CCC
 TCC CAG GAG AGG AAC TTT ACC ACG GCC CCC GGC ATC
 TGC CAC GAG GGC AAG GGC TAC TTC CCC AGA GAG GGC
 GTG TTC GTT TTT AAC GGC ACT AGC TGG TTT ATT ACC
 CAG AGG AAC TTC TTC TCC CCC CAG ATT ATA ACA ACA
 GAT AAC ACT TTC GTG TCC GGC AAC TGC GAT GTT GTG
 ATA GGC ATC ATT AAC AAC ACA GTG TAC GAT CCC CTG
 CAG CCC GAG CTG GAT AGT TTT AAG GAG GAG CTG GAC
 AAG TAT TTT AAG AAC CAC ACT TCC CCC GAT GTA GAC
 CTG GGC GAT ATC AGT GGC ATA AAC GGC AGT GTC GTG
 AAC ATA CAG AAG GAG ATC GAT AGG CTG AAC GAG GTG
 GGC AAG AAC CTG AAG GAG TCA CTG ATC GAT CTG CAG
 GAG CTG GGC AAG TAC GAG CAG TAT ATT AAG TGG CCC

[0188] A representative codon-optimized coding region
 encoding SEQ ID NO:12 according to the "standardized
 optimization" method is presented herein as SEQ ID NO:
 69.

ATG GAT GCC ATG AAG GGC GGC CTG TGC TGT GTG CTG
 CTG CTG TGT GGC GGC GTG TTC GTG AGC CCC AGC GCC
 GGC GGC AGC GGC GAT AGC AGC ATC GGC TAC AGC AAC
 AAC ACC ATC GGC ATC CCC ACC AAC TTC AGC ATC AGC
 ATC ACC ACC GAG GTG ATC CCC GTG AGC ATG GCC AAG
 ACC AGC GTG GAT TGC AAC ATG TAC ATC TGC GGC GAC
 AGC ACC GAG TGC GGC AAC CTG CTG CTG CAG TAC GGC

-continued

AGC TTC TGC ACC CAG CTG AAC CCC GGC CTC AGC GGC
 ATC GCC GGC GAG CAG GAC CCG AAC ACC CCG GAG GTG
 TTC GCC CAG GTG AAG CAG ATG TAC AAG ACC CCC ACC
 CTG AAG TAC TTC GGC GGC TTC AAC TTC AGC CAG ATC
 CTG CCC GAC CCC CTG AAG CCC ACC AAG GGC AGC TTC
 ATC GAG GAT CTG CTG TTC AAC AAG GTG ACC CTG GGC
 GAC GCC GGC TTC ATG AAG CAG TAC GGC GAG TCC CTG
 GGC GAC ATC AAC GGC CCG GAC CTG ATC TGC GGC GAC
 AAG TTC AAC GGC CTG ACC GTG CTG CCC CCC CTG CTG
 ACC GAT GAC ATG ATC GGC GGC TAC AAC GGC GGC CTG
 GTG AGC GGC ACC GGC ACC GGC CCG TGG ACC TTC GGC
 GCC GGC GGC GGC CTG CAG ATC CCC TTC GGC ATG CAG
 ATG GGC TAC CCG TTC AAC GGC ATC GGC GTG ACC CAG
 AAC GTG CTG TAC GAG AAC CAG AAG CAG ATC GCC AAC
 CAG TTC AAC AAG GGC ATC AGC CAG ATC CAG GAG AGC
 CTG ACC ACC ACC AGC ACC GGC CTC GGC AAG CTG CAG
 GAT GTG GTG AAC CAG AAC GGC CAG GGC CTG AAC ACC
 CTG GTG AAG CAG CTG AGC AGC AAC TTC GGC GGC ATC
 AGC AGC CTG GTG AAC GAT ATC CTG AGC GGC CTG GAT
 AAG GTG GAG GGC GAG GTG CAG ATC GAC CCG CTG ATC
 ACC GGC GCG CTG CAG AGC CTG CAG ACC TAC GTG ACC
 CAG CAG CTG ATC CCG GGC GGC GAG ATC CCG GGC AGC
 GCC AAC CTG GGC GGC ACC AAG ATG AGC GAG TGC GTG
 CTG GGC CAG AGC AAG CCG GTG GAT TTC TGC GGC AAG
 GGC TAC CAC CTG ATG AGC TTC CCC CAG GGC GGC CCC
 CAC GGC GTG GTG TTC CTG CAT GTG ACC TAC GTG CCC
 AGC CAG GAG CCG AAC TTC ACC ACC GGC CCC GGC ATC
 TGC CAG GAG GGC AAG GGC TAC TTC CCC CCG GAG GGC
 GTG TTC GTG TTC AAC GGC ACC AGC TGG TTC ATC ACC
 CAG GGC AAC TTC TTC AGC CCC CAG ATC ATC ACC ACC
 GAC AAC ACC TTC GTG AGC GGC AAC TGC GAG GTG GTG
 ATC GGC ATC ATC AAC AAC ACC GTG TAC GAT CCC CTG
 CAG CCC GAG CTG GAT AGC TTC AAG GAG GAG CTG GAC
 AAG TAC TTC AAG AAC CAT ACC AGC CCC GAT GTG GAT
 CTG GGC CAG ATC AGC GGC ATC AAC GGC AGC GTG GTG
 AAC ATC CAG AAG GAG ATC GAT CCC CTG AAC GAG GTG
 GCC AAG AAC CTG AAC GAG AGC CTG ATC GAT CTG CAG
 GAG CTG GGC AAG TAC GAG CAG TAC ATC AAG TGG CCC
 TGG

[0189] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:14 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:14 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:14, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:14 is shown in Table 15.

TABLE 15

AMINO ACID		Number in SEQ ID NO: 14
A	Ala	34
R	Arg	31
C	Cys	0
G	Gly	45
H	His	5
I	Ile	11
L	Leu	26
K	Lys	29
M	Met	7
F	Phe	13
P	Pro	31
S	Ser	35
T	Thr	33
W	Trp	5
Y	Tyr	11
V	Val	11
N	Asn	25
D	Asp	22
Q	Gln	34
E	Glu	14

[0190] Using the amino acid composition shown in Table 15, a human codon-optimized coding region which encodes SEQ ID NO:14 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:14 as follows: the 13 phenylalanine codons are TTC, the 26 leucine codons are CTG, the 11 isoleucine codons are ATC, the 7 methionine codons are ATG, the 11 valine codons are GTG, the 35 serine codons are AGC, the 31 proline codons are CCC, the 33 threonine codons are ACC, the 34 alanine codons are GCC, the 11 tyrosine codons are TAC, the 5 histidine codons are CAC, the 34 glutamine codons are CAG, the 25 asparagine codons are AAC, the 29 lysine codons are AAG, the 22 aspartic acid codons are GAC, the 14 glutamic acid codons are GAG, the 5 tryptophan codons are TGG, the 31 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 45 glycine codons are GGC. The codon-optimized N

coding region designed by this method is presented herein as SEQ ID NO:37.

```

ATGAGCGACAAACGCGCCAGAGCAACAGAGAGCGCCCGAGAAATCAC
CTTCGGCGGCCACCGACAGCAGCAGCAACACAGAAACGCGCGAGAA
ACGGCGCCAGACCCAGCAGAGAGAGCCCGAGGCGTCCCAACACACC
GCCAGCTGGTTACCGCCCTGACCCAGCAGCGCAAGAGAGAGCTAGATT
CCCGAGAGCGCAGGCGCTGCCCTACACACCAACAGCGGCCCGAGAGCC
AGATGCGCTACTACAGAGAGCCACCGAAGAGATGAGAGCGCGGATCGGC
AAGATGAGAGAGCTGAGGCCAGATGCTACTTCTACTACTCGGACCGG
CCCGAGGCCAGCTGCCCTGATCGGCCCAACAGAGAGGACATCTGTGGG
TGCCCAACGAGGCGCGCTGAAACCCCCAGAGACACATCGGACACGAGA
AAACCCAAACAAACAGCGCGCCACCGTCTGCTGACCTGCCCGAGGCCACAC
CCTGCCAAGGGCTTCTACGCGAGGCGAGCAGAGCGCGCAGCCAGGCGCA
CGAGCAGAGCAGCAGCAGAAAGCAGAGGCACACAGAAACAGCAGCCCCC
GCAGCAGCAGAGGCAACAGCGCCCGCAGAGTGGCCAGCGCGCGCGCA
GACCGCGCTGGCCCTGCTGCTGCTGACAGACTGAACACAGCTGGAGAGCA
AGGTGAGCGGCAAGGCGCAGCAGCAGCGCCAGAGCCCTGACCAAGAG
AGGCGCCCGAGAGCGCAGAGAGGCCAGCAGAGAGAGACCGCACCAA
CGCTACAACTGACCCAGCGCTCTGGCAGAGAGAGGCCCGAGCAGACCC
AGGGCAACTTGGCGACAGAGACCTGATCAGCAGAGCGCCAGCATCAG
CACTGGCCCCAGATGCGCCAGTTGCCCCAGCGCGCTCTCTGGG
CATGAGCAGAAATCGCATGAGGTTGACCCCGAGCGGACCTGGCTGACT
ACCAAGCGGCCATCAAGCTGGACGAGAGAGCCCGCTGTTCAAGGACAC
GTGATCTGCTCTAAGCAGCATGAGCGCTGACAGAGCTTCCGCCCCAC
CGAGCCCAAGAGGACAAAGAGAGAGAGACCGAGCGAGGCCCGCTGCG
CCAGAGAGCAGAGAGAGCGCCACCGTGAACCTGCTGCGCCCGCGCGAC
ATGAGCACTTCAGCAGACAGCTGAGAGACAGCATGAGCGCGCGCCAGCGC
CGACAGCACCCAGGCC

```

[0191] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:14 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:14 as follows: about 4 of the 13 phenylalanine codons are TTT, and about 9 of the phenylalanine codons are TTC; about 1 of the 26 leucine codons are TTA, about 6 of the leucine codons are TTG, about 7 of the leucine codons are CTI, about 3 of the leucine codons are CTC, about 5 of the leucine codons are CTA, and about 4 of the leucine codons are CTG; about 7 of the 11 isoleucine codons are ATT, about 3 of the isoleucine codons are ATC, and about 1 of the isoleucine codons are ATA; the 7 methionine codons are ATG; about 4 of the 11 valine codons are GTT, about 4 of the valine codons are GTC, about 1 of the valine codons is GTA, and about 2 of the

valine codons are GTG; about 10 of the 35 serine codons are TCT, about 3 of the serine codons are TCC, about 9 of the serine codons are TCA, about 1 of the serine codons is TCG, about 7 of the serine codons are AGT, and about 5 of the serine codons are AGC; about 10 of the 31 proline codons are CCT, about 9 of the proline codons are CCC, about 10 of the proline codons are CCA, and about 2 of the proline codons are CCG; about 17 of the 33 threonine codons are ACT, about 5 of the threonine codons are ACC, about 11 of the threonine codons are ACA, and about 0 of the threonine codons is ACG; about 14 of the 34 alanine codons are GCT, about 8 of the alanine codons are GCC, about 9 of the alanine codons are GCA, and about 3 of the alanine codons are GCG; about 2 of the 11 tyrosine codons are TAT and about 9 of the tyrosine codons are TAC; about 3 of the 5 histidine codons are CAT and about 2 of the histidine codons are CAC; about 24 of the 34 glutamine codons are CAA and about 10 of the glutamine codons are CAG; about 16 of the 25 asparagine codons are AAT and about 9 of the asparagine codons are AAC; about 20 of the 29 lysine codons are AAA and about 9 of the lysine codons are AAG; about 10 of the 22 aspartic acid codons are GAT and about 12 of the aspartic acid codons are GAC; about 7 of the 14 glutamic acid codons are GAA and about 7 of the glutamic acid codons are GAG; the 5 tryptophan codons are TGG; about 5 of the 31 arginine codons are CGT, about 8 of the arginine codons are CGC, about 6 of the arginine codons are CGA, about 0 of the arginine codons are CCG, about 10 of the arginine codons are AGA, and about 2 of the arginine codons are AGG; and about 10 of the 45 glycine codons are GGT, about 16 of the glycine codons are GGC, about 16 of the glycine codons are GGA, and about 3 of the glycine codons are GGG.

[0192] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0193] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:14, optimized according to codon usage in humans is presented herein as SEQ ID NO:36.

```
ATG TCC GAT AAT GGT CCG CAG TCT AAC CAG AGG TCG
GGG CCA AGA ATC ACA TTC GGG GGC CCA ACA GAG AGT
ACC GAT AAC AAC CAG AAC GGC GGA AGA AAC GGC GCC
AGG CCC AAG CAG CGG AGA CCT CAG GGA TTA CCA AAT
AAT ACC GCA AGC TGG TTC ACA GCC CTG ACC CAG CAT
GGA AAA GAG GAA CTG AGA TTC CTT AGA GGA CAA GGG
GTG CCT ATT AAT ACT AAT AGC GGG CCT GAC GAT CAA
ATT GGC TAT TAT CGA CGT CGC ACT CGC GGT GTT AGA
GGG GGG GAG GGG AAG ATG AAG GAG CTT AGC CCA CGC
TGG TAC TTT TAC TAT CTG GGA ACC GGA CTT GAA GCT
```

-continued

```
AGT CTG CCC TAC GGC GCT AAC AAG GAG GGA ATA GTA
TGG GTC GCC ACG GAA GGT GCG TTG AAT ACT CCG AAA
GAT CAC ATC GGC ACC AGA AAT CTT AAC AAT AAC GCC
GCA ACC GTG CTA CAA TTA CCC CAG GGA ACT ACT CTG
CCG AAG GGG TCT TAT GCG GAG GGA AGC GCG GGC GGC
TCA CAA GCC AGT TCA CGC TCC AGC TCC GCG TCG AGG
GGT AAT TCC CGA AAC AGC ACC CGG GGA TCA TCT AGG
GGA AAC TCT CCC GCG CGG ATG CGC TCA GGC GGC GGC
GAA ACA GCT CTG GCT CTG CTA TTG CTG GAC CGG CTC
AAC CAG CTC GAG TCC AAA GTC TCT GGT AAA GGT CAG
CAG CAG CAG GGT CAA ACA GTG ACC AAA AAA AGT GCA
GCC GAG GCC AGC AAG AAA CCA CGC CAG AAA CGT AGG
GCC ACA AAG CAA TAC AAT GTG ACC CAA GCC TTT GAG
AAG CGG GGG CGG GAA CAG ACA CAG CGC AAT TTC GGC
GAT CAA GAT TTG ATA CGA CAG GCG ACT GAC TAC AAA
CAC TGG CGC CAG ATC GCT CAG TTT GCA CCT AGC GCC
TCC GCT TTT TTT GGC ATG AGT CGG ATT GGC ATG GAG
GTG ACA CCA TCA GGT ACT TGG TTA ACG TAC CAC GGG
GCA ATT AAA CTT GAT GAT AAA GAT CCC CAG TTT AAG
GAC AAC GTT ATC CTC CTG AAT AAG CAT ATT GAC GCC
TAT AAG ACC TTC CCC CCA ACC GAA CCA AAG AAG GAC
AAG AAG AAG AAG ACA GAC GAG GCA CAG CCT CTC GCC
CAG AGG CAG AAA AAG CAG CCT ACT GTG ACC CTT CTG
CCC GCT GCA GAG ATG GAT GAC TTT TCC CGC CAA CTC
CAG AAC TCT ATG AGT GGG GCT TCC GCT GAC TCT AGG
CAG GCC TGA
```

[0194] Another representative codon-optimized coding region encoding SEQ ID NO:14 is presented herein as SEQ ID NO:63. SEQ ID NO:14 is encoded by nucleotides 7 to 1275 of SEQ ID NO:63.

```
GTGCGATGAGCGACACGAGGCCGACGACACAGAGAGGCCCCCAG
AATCACCTTTGGCGGCTCTACGACAGCAGCAGCAACACAGAGCGCG
CGAGAAACGGCGCAGACCCAGCAGAGAGGACGCCCGGCTCTGCCAAC
AACAGCCGCCAGCTGGTTACCGGCTCTACCGACAGCGCAAGAGAGCTC
GAGATTCCCCAGAGCGCCAGGGCGTGGCCATCAATCAACAGCGGCCGAG
ACGATCAGATCGCTACTACCGAGGCGCAGCAGAGAGTGTAGAGCGGCG
GAGCGCAGATGAGGAGCTGAGCCCCCGTGGTACTTCTACTACTCTGGG
CACCGGCCCTGAGGCGCAGCTGCTGCTACCGGCGCAACAGAGAGGAGCTG
TGTGGGTGCCACGAGGGGCGCCCTGATACCCCAAGGACCACTCGCG
```


-continued

TCAGCAGACAGCTTCGAGAACAGCATGAGCGGCCAGCGCCGACAGCACAC
CAGGCC

[0197] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:16 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:16 as follows: about 5 of the 12 phenylalanine codons are TTT, and about 7 of the phenylalanine codons are TTC; about 3 of the 26 leucine codons are TTA, about 3 of the leucine codons are TTG, about 3 of the leucine codons are CTT, about 5 of the leucine codons are CTC, about 2 of the leucine codons are CTA, and about 10 of the leucine codons are CTG; about 4 of the 11 isoleucine codons are ATT, about 5 of the isoleucine codons are ATC, and about 2 of the isoleucine codons are ATA; the 7 methionine codons are ATG; about 2 of the 11 valine codons are GTT, about 3 of the valine codons are GTC, about 1 of the valine codons is GTA, and about 5 of the valine codons are GTG; about 6 of the 35 serine codons are TCT, about 8 of the serine codons are TCC, about 5 of the serine codons are TCA, about 2 of the serine codons are TCG, about 6 of the serine codons are AGT, and about 8 of the serine codons are AGC; about 8 of the 28 proline codons are CCT, about 9 of the proline codons are CCC, about 8 of the proline codons are CCA, and about 3 of the proline codons are CCG; about 7 of the 30 threonine codons are ACT, about 11 of the threonine codons are ACC, about 9 of the threonine codons are ACA, and about 3 of the threonine codons are ACG; about 9 of the 33 alanine codons are GCT, about 13 of the alanine codons are GCC, about 7 of the alanine codons are GCA, and about 4 of the alanine codons are GCG; about 5 of the 11 tyrosine codons are TAT and about 6 of the tyrosine codons are TAC; about 2 of the 5 histidine codons are CAT and about 3 of the histidine codons are CAC; about 9 of the 33 glutamine codons are CAA and about 24 of the glutamine codons are CAG; about 12 of the 25 asparagine codons are AAT and about 13 of the asparagine codons are AAC; about 9 of the 22 lysine codons are AAA and about 13 of the lysine codons are AAG; about 9 of the 20 aspartic acid codons are GAT and about 11 of the aspartic acid codons are GAC; about 5 of the 12 glutamic acid codons are GAA and about 7 of the glutamic acid codons are GAG; the 5 tryptophan codons are TGG; about 3 of the 31 arginine codons are CGT, about 6 of the arginine codons are CGC, about 3 of the arginine codons are CGA, about 6 of the arginine codons are CGG, about 7 of the arginine codons are AGA, and about 6 of the arginine codons are AGG; and about 7 of the 45 glycine codons are GGT, about 15 of the glycine codons are GGC, about 12 of the glycine codons are GGA, and about 11 of the glycine codons are GGG.

[0198] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one

codon encoding a given amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0199] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:16, optimized according to codon usage in humans is presented herein as SEQ ID NO:38.

ATG AGT GAT AAT AAC GGC CCC CAG TCT AAC CAG AGG AGC
GCA CCG CGG ATC ACG TTC GGT GGC CCA ACC GAC TCA
ACA GAC AAT AAT CAG AAC GGA GGA CGC AAT GGT GCA
CGT CCT AAG CAG AGA CGC CCC CAA GGG CTG CCT AAT
AAT ACA GCA AGT TGG TTT ACC GCA CTC ACA CAA CAT
GGA AAG GAA GAG TTG CGG TTC CCC CGC GGC CAG GGC
GTG CCC ATC AAC ACA AAT AGC GGA CCC GAC GAT CAG
ATC GGA TAT TAC CAG AGA GCT ACA AGG AGA GTT CGC
GGC GGG GAT GGC AAG ATG AAG GAG CTA TCA CCA CGA
TGG TAC TTC TAT TAC CTC GGG ACA GGC CCA GAG CGC
TCG CTA CCA TAC GGG GCC AAC AAG GAG GGT ATT GTC
TGG GTC GCT ACC GAA GGG GCC CTG AAT ACA CCT AAA
GAC CAC ATA GGT ACC AGA AAT CCC AAC AAT AAC GCC
CGC AGC GTG TTA CAG CTT CCT CAG GGA ACG ACC CTT
CCA AAA GGG TTT TAC GCC GAA GGA TCT CGG GGA GGG
TCA CAG GCT AGC TCC CGT AGC TCC TCA AAG TCC AGG
GGG AAT TCT AGA AAC AGT ACA CCC GGC TCT AGC CGT
GGT AAC TCC CCA GCT CGC ATG GCA TCC GCC GGA GGG
GAA ACC GCT CTG GCT CTG CTC CTA TTA GAT CGG TTG
AAC CAA CTG GAA TCG AAG GTA TCG AGC AAG GAG TCG
CAG CAG CAA GGC CAG ACT GTG ACT AAG AAG TCC GCG
GCC GAG GCC AGT AAG AAA CCC CGC CAG AAA CGA ACT
GCC ACC AAA CAG TAT AAT GTG ACA CAG GCC TTC GGC
AGA GGG GGT CCA GAG CAG ACC CAA GGC AAC TCT GGG
GAT CAG GAC CTG ATT CGG CAG GGT ACC GAC TAT AAG
CAC TGG CCG CAA ATT GCT CAG TTT GCT CCC AGT GGG
AGT GCC TTC TTC GGC ATG TCT AAG AAT GGG AAG GAG
GTT ACT CCT AGC GGC ACT TGG CTT ACT TAT CAC GGA
GCC ATC AAA CTC GAT AAG AAG CCA CAG CTT AAG
GAT AAC GTG ATT CTG CAG AAC AAA CAT ATA GAC GCG
TAC CCT CTC CCG CAA AAG CAG AAA AAA CAG CTT ACC
GTC ACG TTA CTG CTT GCC CCA GAC ATG GAC GAC TTT
TCT AGA CAG TTG CAA AAC AGC ATG TCA GGC GCA TCC
GCC GAT AGC ACT CAA GCT TGA

[0200] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:19 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:19 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:19, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:19 is shown in Table 17.

TABLE 17

AMINO ACID	Number in SEQ ID NO: 19
A	Ala 19
R	Arg 15
C	Cys 3
G	Gly 15
H	His 3
I	Ile 18
L	Leu 31
K	Lys 6
M	Met 7
F	Phe 11
P	Pro 6
S	Ser 11
T	Thr 13
W	Trp 7
Y	Tyr 9
V	Val 16
N	Asn 13
D	Asp 6
Q	Gln 5
E	Glu 7

[0201] Using the amino acid composition shown in Table 17, a human codon-optimized coding region which encodes SEQ ID NO:19 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:19 as follows: the 11 phenylalanine codons are TTC, the 31 leucine codons are CTG, the 18 isoleucine codons are ATC, the 7 methionine codons are ATG, the 16 valine codons are GTG, the 11 serine codons are AGC, the 6 proline codons are CCC, the 13 threonine codons are ACC, the 19 alanine codons are GCC, the 19 tyrosine codons are TAC, the 3 histidine codons are CAC, the 5 glutamine codons are CAG, the 13 asparagine codons are AAC, the 6 lysine codons are AAG, the 6 aspartic acid codons are GAC, the 7 glutamic acid codons are GAG, the 3 cysteine codons are TGC, the 7 tryptophan codons are TGG, the 15 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 43 glycine codons are GGC. The codon-optimized M coding region designed by this method is presented herein as SEQ ID NO:41.

ATGCGCCAGAACGGCACCATCACCGTGGAGAGCTGAAGCAGCTCTGGA

GAGTGGAACTGGTGATGGCTTCCTGCTTGGCTGGATCATGCTGC

TGCAGTTGGCTACAGCAACAGAACAGATCTCTCATCATCATGCACTG

—continued

GTGTTCTCTGTGGCTGCTGTGTGGCCCTGACCTGGCTGCTGTGCTGGC

COCCTGTATCAGATCAACTGGGTGACCGCGGCGATGCGATCGCATGG

CTGTGATCTGTGGCTGATGTGGCTGAGCTACCTTGTGGCCAGCTTCAGA

CTTTTCGCCGAACCAAGACATGTGGAGCTTCAACCCGAGACCAACT

CTGTGTGAACGTGCCCCGTGAGGACACCATGTGACCAAGACCTCTGATGG

AGAGCGAGCTGTGTATGGCGCCCTGTATCATCAGAGCCACTGTGAATG

CGCGGCCACCCCTGGGCAGATGCGACATAGGACCTGCCCAAGGAGAT

CACCTGTGCCACCGACAGACCTGTGAGCTACTCAAGCTGGCGCCAGCC

AGAGAGTGGGCACCGACACCGCCTTTCGCCGCTCTACACAGATACAGATG

GGCAACTACAGCTGAAACACCGACACCGCGCGGCAAGGACACATCGC

CTCTGTGTCAG

[0202] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:19 can be designed by the "full optimization" method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:19 as follows: about 5 of the 11 phenylalanine codons are TTT, and about 6 of the phenylalanine codons are TTC; about 3 of the 31 leucine codons are TTA, about 4 of the leucine codons are TTG, about 4 of the leucine codons are CTT, about 6 of the leucine codons are CTC, about 2 of the leucine codons are CTA, and about 12 of the leucine codons are CTG; about 6 of the 18 isoleucine codons are ATT, about 9 of the isoleucine codons are ATC, and about 3 of the isoleucine codons are ATA; the 7 methionine codons are ATG; about 3 of the 16 valine codons are GTT, about 4 of the valine codons are GTC, about 2 of the valine codons are GTA, and about 7 of the valine codons are GTG; about 2 of the 11 serine codons are TCT, about 2 of the serine codons are TCC, about 2 of the serine codons are TCA, about 1 of the serine codons is TCG, about 1 of the serine codons is AGT, and about 3 of the serine codons are AGC; about 2 of the 6 proline codons are CCT, about 2 of the proline codons are CCC, about 1 of the proline codons is CCA, and about 1 of the proline codons is CCG; about 3 of the 13 threonine codons are ACT, about 5 of the threonine codons are ACC, about 4 of the threonine codons are ACA, and about 1 of the threonine codons is ACG; about 5 of the 19 alanine codons are GAT, about 8 of the alanine codons are GCC, about 4 of the alanine codons are GCA, and about 2 of the alanine codons are GCG; about 4 of the 9 tyrosine codons are TAT and about 5 of the tyrosine codons are TAC; about 1 of the 3 histidine codons is CAT and about 2 of the histidine codons are CAC; about 1 of the 5 glutamine codons is CAA and about 4 of the glutamine codons are CAG; about 6 of the 13 asparagine codons are AAT and about 7 of the asparagine codons are AAC; about 3 of the 6 lysine codons are AAA and about 3 of the lysine codons are AAG; about 3 of the 6 aspartic acid codons are GAT and about 3 of the aspartic acid codons are GAC; about 3 of the 7 glutamic acid codons are GAA and about 4 of the glutamic acid codons are GAG; about 1 of the 3 cysteine codons is TGT and about 2 of the cysteine codons are TGC; the 7 tryptophan codons are TGG; about 1 of the

15 arginine codons is CGT, about 3 of the arginine codons are CGC, about 2 of the arginine codons are CGA, about 3 of the arginine codons are CGG, about 3 of the arginine codons are AGA, and about 3 of the arginine codons are AGG; and about 2 of the 15 glycine codons are GGT, about 5 of the glycine codons are GGC, about 4 of the glycine codons are GGA, and about 4 of the glycine codons are GGG.

[0203] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a give amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0204] A representative "fully optimized" codon-optimized coding region encoding SEQ ID NO:19, optimized according to codon usage in humans is presented herein as SEQ ID NO:40.

ATG GCT GAC AAC GGC ACC ATA ACC CTC GAG GAG CTT
AAA CAG TTA TTA GAA CAA TGG AAT TTG GTG ATA GGA
TTC CTC TTT CTC GCA TGG ATC ATG TTT CTT CAG TTC
GCC TAT TCT AAC CCG AAT AGG TTT TTG TAC ATT ATC
AAG CTG GTC TTC CTT TGG CTG CTC TGG CCC GTA ACA
CTA GCC TGT TTT GTT TTG GCG GCC GTG TAT CGG ATC
AAT TGG GTG ACA GGT GGC ATT GCT ATT GCG ATG GCT
TGC ATC GTG GGG CTG ATG TGG CTG TCG TAT TTC GTT
GCC TCA TTC CCG GTT TTT GCC CGA ACA AGG AGT ATG
TGG TCT TTT AAC CCG GAG ACC AAT ATT CTC ATC AAT
GTG CCT TTA CGC GGC ACT ATG GTG ACC CGC CTT CTA
ATG GAA TCC GAG CTG GTA ATT GGC GCA CTC ATC ATA
AGG GGG CAC CTC AGA ATT GCC GGG CAC CCA CTT GGG
AGA TGC GAC ATC AAG GAT CTC CGG AAG GAA ATT ACT
GTT GCA ACT TCA CGA CGG CTG AGC TAT TAC AAA CTG
GGA GCT AGC CAG AGA GTG GGT ACC GAC TAC GGC TTC
GCT GCC TAC AAC CGC TAC COT ATC GGA AAT TAC AAA
CTC AAC ACA GAT CAT CAG GGA AGC AAT GAT AAC ATC
GCC CTC CTG CTC CAG TGA

[0205] In certain embodiments described herein, a codon-optimized coding region encoding SEQ ID NO:21 is optimized according to codon usage in humans (*Homo sapiens*). Alternatively, a codon-optimized coding region encoding SEQ ID NO:21 may be optimized according to codon usage in any plant, animal, or microbial species. Codon-optimized coding regions encoding SEQ ID NO:21, optimized according to codon usage in humans are designed as follows. The amino acid composition of SEQ ID NO:21 is shown in Table 18.

TABLE 18

AMINO ACID		Number in SEQ ID NO: 21
A	Ala	4
R	Arg	2
C	Cys	3
G	Gly	2
H	His	0
I	Ile	3
L	Leu	14
K	Lys	2
M	Met	1
F	Phe	4
P	Pro	2
S	Ser	7
T	Thr	5
W	Trp	0
Y	Tyr	4
V	Val	14
N	Asn	1
D	Asp	1
Q	Gln	0
E	Glu	3

[0206] Using the amino acid composition shown in Table 18, a human codon-optimized coding region which encodes SEQ ID NO:21 can be designed by any of the methods discussed herein. For "uniform" optimization, each amino acid is assigned the most frequent codon used in the human genome for that amino acid. According to this method, codons are assigned to the coding region encoding SEQ ID NO:21 as follows: the 4 phenylalanine codons are TTC, the 14 leucine codons are CTG, the 18 isoleucine codons are 3, the 1 methionine codon is ATG, the 14 valine codons are GTG, the 7 serine codons are AGC, the 2 proline codons are CCC, the 5 threonine codons are ACC, the 4 alanine codons are GCC, the 4 tyrosine codons are TAC, the 5 asparagine codons are AAC, the 2 lysine codons are AAG, the 1 aspartic acid codon is GAC, the 3 glutamic acid codons are GAG, the 3 cysteine codons are TGC, the 1 tryptophan codon is TGG, the 2 arginine codons are CGG, AGA, or AGG (the frequencies of usage of these three codons in the human genome are not significantly different), and the 2 glycine codons are GGC. The codon-optimized E coding region designed by this method is presented herein as SEQ ID NO:43.

ATG TAC ACC TTC GTG AGC GAG ACC GGC ACC CTG
ATC CTG AAC AGC GTG CTG CTG TTC CTG GGC TTC GTG
GTG TTC CTG CTG GTG ACC CTG GCC ATC CTG ACC GCC
CTG CGG CTG TGC GCC TAC TGC TGC AAC ATC GTG AAC
GTG ACC CTG GTG AAG CCC ACC GTG TAC GTG TAC AGC
CGG GTG AAG AAC CTG AAC AGC GAG GGC GTG CCC
GAC CTG CTG GTG TGA

[0207] Alternatively, a human codon-optimized coding region which encodes SEQ ID NO:21 can be designed by an optimization method, where each amino acid is assigned codons based on the frequency of usage in the human genome. These frequencies are shown in Table 4 above. Using this latter method, codons are assigned to the coding region encoding SEQ ID NO:21 as follows: about 1 of the

4 phenylalanine codons are TTT, and about 3 of the phenylalanine codons are TTC; about 2 of the 14 leucine codons are TTA, about 2 of the leucine codons are TTG, about 6 of the leucine codons are CTT, about 0 of the leucine codons are CTC, about 2 of the leucine codons are CTA, and about 2 of the leucine codons are CTG; about 1 of the 3 isoleucine codons are ATT, about 1 of the isoleucine codons are ATC, and about 1 of the isoleucine codons are ATA; the 1 methionine codons are ATG; about 6 of the 14 valine codons are GTT, about 3 of the valine codons are GTC, about 3 of the valine codons are GTA, and about 2 of the valine codons are GTG; about 2 of the 7 serine codons are TCT, about 0 of the serine codons are TCC, about 1 of the serine codons are TCA, about 2 of the serine codons are TCG, about 1 of the serine codons are AGT, and about 1 of the serine codons are AGC; about 1 of the 2 proline codons are CCT, about 0 of the proline codons are CCC, about 1 of the proline codons is CCA, and about 0 of the proline codons is CCG; about 1 of the 5 threonine codons are ACT, about 0 of the threonine codons are ACC, about 2 of the threonine codons are ACA, and about 2 of the threonine codons are ACG; about 1 of the 4 alanine codons are GCT, about 1 of the alanine codons are GCC, about 0 of the alanine codons are GCA, and about 2 of the alanine codons are GCG; about 0 of the 4 tyrosine codons are TAT and about 4 of the tyrosine codons are TAC; about 3 of the 5 asparagine codons are AAT and about 2 of the asparagine codons are AAC; about 2 of the 2 lysine codons are AAA and about 0 of the lysine codons are AAG; about 1 of the 1 aspartic acid codons are GAT and about 0 of the aspartic acid codons are GAC; about 3 of the 3 glutamic acid codons are GAA and about 0 of the glutamic acid codons are GAG; about 1 of the 3 cysteine codons is TGT and about 2 of the cysteine codons are TGC; about 1 of the 2 arginine codons is CGT, about 0 of the arginine codons are CGC, about 1 of the arginine codons are CGA, about 0 of the arginine codons are CGG, about 0 of the arginine codons are AGA, and about 0 of the arginine codons are AGG; and about 1 of the 2 glycine codons are GGT, about 0 of the glycine codons are GGC, about 1 of the glycine codons are GGA, and about 0 of the glycine codons are GGG.

[0208] As described above, the term "about" means that the number of amino acids encoded by a certain codon may be one more or one less than the number given. It would be understood by those of ordinary skill in the art that the total number of any amino acid in the polypeptide sequence must remain constant, therefore, if there is one "more" of one codon encoding a given amino acid, there would have to be one "less" of another codon encoding that same amino acid.

[0209] A representative fully codon-optimized coding region encoding SEQ ID NO:21, optimized according to codon usage in humans is presented herein as SEQ ID NO:42.

ATG TAC AGC TTT GTG TCT GAA GAA ACA GGA AGG TTG
ATA GTT AAT AGT GTT TTG CTT TTC TTA GCG TTC GTA
GTC TTC CTT CTT GTC ACA CTT GCC ATT TTA ACT GCG
CTT COT CTA TGC GCT TAC TGT TGC AAT ATC GTA AAC
GTG TCG CTT GTT AAA CCA AGC GTT TAC GTA TAC TCG

—continued

GGA GTT AAA AAC CTG AAT TCT TTA GAA GGT GTT CCT
GAT CTG CTA GTC TAA

[0210] Another representative codon-optimized coding region encoding SEQ ID NO:21 is presented herein as SEQ ID NO:48.

ATG TAT AGT TTT GTG AGT GAG GAG ACG GCC ACC CTG
ATT GTC AAC TCA GTG CTG CTG TTC CTG GCC TTT GTT
GTC TTC CTG CTG GTA ACT CTG GCC ATC CTG ACT GCC
CTG AGA GTG TGC GCC TAC TGC TGC AAC ATC GTG AAC
GTC TCT CTG GTA AAG CCC ACA GTT TAC GTG TAT TCT
AGG GTG AGC AAG CTA AAC TCC AGC GAG GGC GTT CCC
GAT CTG CTG GTA TAA

[0211] Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence using the "uniform optimization," "full optimization," "minimal optimization," or other optimization methods, can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNASTar, Inc., Madison, WI, the backtranslation function in the VectorNTI Suite, available from Informax, Inc., Bethesda, Md., and the "backtranslate" function in the GCG—Wisconsin Package, available from Accelrys, Inc., San Diego, Calif. In addition, various resources are publicly available to codon-optimize coding region sequences. For example, the "backtranslation" function found at <http://www.entelchion.com/eng/backtranslation.html> (visited Jul. 9, 2002), and the "backtranseq" function available at <http://bioinfo.pbi.nrc.ca:8090/EMBOSS/index.html> (visited Oct. 15, 2002). Constructing a rudimentary algorithm to assign codons based on a given frequency can also easily be accomplished with basic mathematical functions by one of ordinary skill in the art.

[0212] A number of options are available for synthesizing codon-optimized coding regions designed by any of the methods described above, using standard and routine molecular biological manipulations well known to those of ordinary skill in the art. In one approach, a series of complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of the desired sequence are synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends, e.g., each oligonucleotide in the pair is synthesized to extend 3, 4, 5, 6, 7, 8, 9, 10, or more bases beyond the region that is complementary to the other oligonucleotide in the pair. The single-stranded ends of each pair of oligonucleotides is designed to anneal with the single-stranded end of another pair of oligonucleotides. The oligonucleotide pairs are allowed to anneal, and approximately five to six of these double-stranded fragments are

then allowed to anneal together via the cohesive single stranded ends, and then they ligated together and cloned into a standard bacterial cloning vector, for example, a TOPO® vector available from Invitrogen Corporation, Carlsbad, Calif. The construct is then sequenced by standard methods. Several of these constructs consisting of 5 to 6 fragments of 80 to 90 base pair fragments ligated together, i.e., fragments of about 500 base pairs, are prepared, such that the entire desired sequence is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced. Additional methods would be immediately apparent to the skilled artisan. In addition, gene synthesis is readily available commercially.

[0213] The codon-optimized coding regions can be versions encoding any gene products from any strain, derivative, or variant of SARS-CoV, or fragments, variants, or derivatives of such gene products. For example, nucleic acid fragments of codon-optimized coding regions encoding the S, N, E or M polypeptides, or fragments, variants or derivatives thereof. Codon-optimized coding regions encoding other SARS-CoV polypeptides or fragments, variants, or derivatives thereof (e.g., those encoding certain predicted open reading frames in the SARS-CoV genome), are included within the present invention. Additional, non-codon-optimized polynucleotides encoding SARS-CoV polypeptides or other polypeptides may be included as well.

Compositions and Methods

[0214] In certain embodiments, the present invention is directed to compositions and methods of raising a detectable immune in a vertebrate by administering *in vivo*, into a tissue of a vertebrate, one or more polynucleotides comprising at least one wild-type coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof, and/or at least one codon-optimized coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof. In addition, the present invention is directed to compositions and methods of raising a detectable immune response in a vertebrate by administering to the vertebrate a composition comprising one or more polynucleotides as described herein, and at least one isolated SARS-CoV component, or isolated polypeptide. The SARS-CoV component may be inactivated virus, attenuated virus, a viral vector expressing an isolated SARS-CoV polypeptide, or a SARS-CoV virus protein, fragment, variant or derivative thereof.

[0215] The polynucleotides comprising at least one coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof, and/or at least one codon-optimized coding region encoding a SARS-CoV polypeptide may be administered either prior to, at the same time (simultaneously), or subsequent to the administration of the SARS-CoV component, or isolated polypeptide.

[0216] The SARS-CoV component, or isolated polypeptide in combination with polynucleotides comprising at least one coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof, and/or at least one codon-optimized coding region encoding a SARS-CoV polypeptide compositions may be referred to as "combinatorial polynucleotide vaccine compositions" or "single formulation heterologous prime-boost vaccine compositions."

[0217] The isolated SARS-CoV polypeptides of the invention may be in any form, and are generated using techniques well known in the art. Examples include isolated SARS-CoV proteins produced recombinantly, isolated SARS-CoV proteins directly purified from their natural milieu, recombinant (non-SARS-COV) virus vectors expressing an isolated SARS-CoV protein, or proteins delivered in the form of an inactivated SARS-CoV vaccine, such as conventional vaccines.

[0218] When utilized, an isolated SARS-CoV component, or polypeptide or fragment, variant or derivative thereof is administered in an immunologically effective amount. Canine coronavirus, known to infect swine, turkeys, mice, calves, dogs, cats, rodents, avians and humans, may be administered as a live viral vector vaccine at a dose rate per dog of 10^4 - 10^8 pfu, or as a typical subunit vaccine at 10 ug- 1 mg of polypeptide, according to U.S. Pat. No. 5,661,006, incorporated by reference herein in its entirety. Similarly, Bovine coronavirus is administered to animals in an antigen vaccine composition at dose of about 1 to about 100 micrograms of subunit antigen, according to U.S. Pat. No. 5,369,026, incorporated by reference herein in its entirety. The effective amount of SARS-CoV component or isolated polypeptide, and polynucleotides as described herein are determinable by one of ordinary skill in the art based upon several factors, including the antigen being expressed, the age and weight of the subject, and the precise condition requiring treatment and its severity, and route of administration.

[0219] In the instant invention, the combination of conventional antigen vaccine compositions with the polynucleotides comprising at least one coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof, and/or at least one codon-optimized coding region encoding a SARS-CoV polypeptide compositions provides for therapeutically beneficial effects at dose sparing concentrations. For example, immunological responses sufficient for a therapeutically beneficial effect in patients predetermined for an approved commercial product, such as for the typical animal coronavirus products described above, may be attained by using less of the product when supplemented or enhanced with the appropriate amount of polynucleotides comprising at least one coding region encoding a SARS-CoV or codon-optimized nucleic acid. Thus, dose sparing is contemplated by administration of conventional coronavirus vaccines administered in combination with the nucleic acids of the invention.

[0220] In particular, the dose of an antigen SARS-CoV vaccine may be reduced by at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70% when administered in combination with the nucleic acid compositions of the invention.

[0221] Similarly, a desirable level of an immunological response afforded by a DNA-based pharmaceutical alone may be attained with less DNA by including an aliquot of antigen SARS-CoV vaccine. Further, using a combination of conventional and DNA-based pharmaceuticals may allow both materials to be used in lesser amounts, while still affording the desired level of immune response arising from administration of either component alone in higher amounts (e.g., one may use less of either immunological product when they are used in combination). This may be manifest

not only by using lower amounts of materials being delivered at any time, but also to leads to reducing the number of administrations in a vaccination regime (e.g., 2 versus 3 or 4 injections), and/or to reducing the kinetics of the immunological response (e.g., desired response levels are attained in 3 weeks instead of 6 weeks after immunization).

[0222] In particular, the dose of DNA-based pharmaceuticals, may be reduced by at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60% or at least 70% when administered in combination with antigen SARS-CoV vaccines.

[0223] Determining the precise amounts of DNA based pharmaceutical and SARS-CoV antigen is based on a number of factors as described above, and is readily determined by one of ordinary skill in the art.

[0224] In addition to dose sparing, the claimed combinatorial compositions provide for a broadening of the immune response and/or enhanced beneficial immune responses. Such broadened or enhanced immune responses are achieved by: adding DNA to enhance cellular responses to a conventional vaccine; adding a conventional vaccine to a DNA pharmaceutical to enhance humoral response; using a combination that induces additional epitopes (both humoral and/or cellular) to be recognized and/or responded to in a more desirable way (epitope broadening); employing a DNA-conventional vaccine combination designed for a particular desired spectrum of immunological responses; and/or obtaining a desirable spectrum by using higher amounts of either component. The broadened immune response is measurable by one of ordinary skill in the art by standard immunological assays specific for the desirable response spectrum.

[0225] Both broadening and dose sparing may be obtained simultaneously.

[0226] In addition, the present invention is directed to compositions and methods of raising a detectable immune response in a vertebrate by administering to the vertebrate a composition comprising one or more SARS-CoV polynucleotides as described herein. The compositions of the invention may comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10 polynucleotides, as described herein, encoding different SARS-CoV polypeptides or fragments, variants or derivatives thereof in the same composition.

[0227] The coding regions encoding SARS-CoV polypeptides or fragments, variants, or derivatives thereof may be codon optimized for a particular vertebrate. Codon optimization is carried out by the methods described herein; for example, in certain embodiments codon-optimized coding regions encoding polypeptides of SARS-CoV, or nucleic acid fragments of such coding regions encoding fragments, variants, or derivatives thereof are optimized according to the codon usage of the particular vertebrate. The polynucleotides of the invention are incorporated into the cells of the vertebrate in vivo, and an immunologically effective amount of a SARS-CoV polypeptide or a fragment, variant, or derivative thereof is produced in vivo. The coding regions encoding a SARS-CoV polypeptide or a fragment, variant, or derivative thereof may be codon optimized for mammals, e.g., humans, apes, monkeys (e.g., owl, squirrel, cebus, rhesus, African green, patas, cynomolgus, and cercopithecus), orangutans, baboons, gibbons, and chimpanzees, dogs, wolves, cats, lions, and tigers, horses, donkeys, zebras, cows, pigs, sheep, deer, giraffes, bears, rabbits, mice, ferrets, seals, whales; birds, e.g., ducks, geese, terns, shearwaters, gulls, turkeys, chickens, quail, pheasants, geese, starlings and budgerigars; or other vertebrates.

[0228] In particular, the present invention relates to codon-optimized coding regions encoding polypeptides of SARS-CoV, or fragments, variants, or derivatives thereof, or nucleic acid fragments of such coding regions or fragments, variants, or derivatives thereof, which have been optimized according to human codon usage. For example, human codon-optimized coding regions encoding polypeptides of SARS-CoV, or fragments, variants, or derivatives thereof are prepared by substituting one or more codons preferred for use in human genes for the codons naturally used in the DNA sequence encoding the SARS-CoV polypeptide or a fragment, variant, or derivative thereof. Also provided are polynucleotides, vectors, and other expression constructs comprising wild-type coding regions or codon-optimized coding regions encoding polypeptides of SARS-CoV, or nucleic acid fragments of such wild-type coding regions or codon-optimized coding regions including variants, or derivatives thereof. Also provided are pharmaceutical compositions comprising polynucleotides, vectors, and other expression constructs comprising wild-type coding regions or codon-optimized coding regions encoding polypeptides of SARS-CoV, or nucleic acid fragments of such coding regions encoding variants, or derivatives thereof; and various methods of using such polynucleotides, vectors and other expression constructs. Coding regions encoding SARS-CoV polypeptides may be uniformly optimized, fully optimized, or minimally optimized, or otherwise optimized, as described herein.

[0229] The present invention is further directed towards polynucleotides comprising coding regions or codon-optimized coding regions encoding polypeptides of SARS-CoV antigens, for example, (predicted ORF's), optionally in conjunction with other antigens. The invention is also directed to polynucleotides comprising nucleic acid fragments or codon-optimized nucleic acid fragments encoding fragments, variants and derivatives of these polypeptides.

[0230] In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment, where the nucleic acid fragment is a fragment of a coding region or a codon optimized coding region encoding a polypeptide at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a SARS-CoV polypeptide, e.g., S, N, E or M, and where the nucleic acid fragment is a variant of a coding region or a codon optimized coding region encoding an SARS-CoV polypeptide, e.g., S, N, E or M. The human codon-optimized coding region can be optimized for any vertebrate species and by any of the methods described herein.

[0231] As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a

subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (*Comp. App. Biosci.* 6:237-245 (1990)). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is expressed as percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining, Penalty=30 Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

Isolated SARS-CoV Polypeptides

[0232] The present invention is further drawn to compositions which include at least one polynucleotide comprising one or more nucleic acid fragments, where each nucleic acid fragment is a fragment of a coding region or a codon-optimized coding region operably encoding an SARS-CoV polypeptide or fragment, variant, or derivative thereof; together with and one or more isolated SARS-CoV, components, polypeptides or fragments, variants or derivatives thereof, i.e., "combinatorial polynucleotide vaccine compositions" or "single formulation heterologous prime-boost vaccine compositions." The isolated SARS-CoV polypeptides of the invention may be in any form, and are generated using techniques well known in the art. Examples include isolated SARS-CoV proteins produced recombinantly, isolated SARS-CoV proteins directly purified from their natural milieu, and recombinant (non-SARS-CoV) virus vectors expressing an isolated SARS-CoV protein.

[0233] Similarly, the isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof to be delivered (either a recombinant protein, a purified subunit, or viral vector expressing an isolated SARS-CoV polypeptide) may be any isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof, including but not limited to the S, S1, S2, N, E or M proteins or fragments, variants or derivatives thereof. Fragments include, but are not limited to the soluble portion of the S protein and the S1 and S2 domains of the S protein. In certain embodiments, a derivative protein may be a fusion protein. It should be noted that any isolated SARS-CoV polypeptide or fragment, variant, or derivative thereof described herein may be combined in a composition with any polynucleotide comprising a nucleic acid fragment, where the nucleic acid fragment is a fragment of a coding region or a codon-optimized coding region operably encoding a SARS-CoV polypeptide or fragment, variant, or derivative thereof. The proteins may be different, the same, or may be combined in any combination of one or more isolated SARS-CoV proteins and one or more polynucleotides.

[0234] In certain embodiments, the isolated SARS-CoV polypeptides, or fragments, derivatives or variants thereof may be fused to or conjugated to a second isolated SARS-CoV polypeptide, or fragment, derivative or variant thereof, or may be fused to other heterologous proteins, including for example, hepatitis B proteins including, but not limited to the hepatitis B core antigen (HBcAg), or those derived from diphtheria or tetanus. The second isolated SARS-CoV polypeptide or other heterologous protein may act as a

"carrier" that potentiates the immunogenicity of the SARS-CoV polypeptide or a fragment, variant, or derivative thereof to which it is attached. Hepatitis B virus proteins and fragments and variants thereof useful as carriers within the scope of the invention are disclosed in U.S. Pat. No. 6,231, 864 and U.S. Pat. No. 5,143,726, incorporated by reference in their entireties. Polynucleotides comprising coding regions encoding said fused or conjugated proteins are also within the scope of the invention.

Methods and Administration

[0235] The present invention also provides methods for delivering a SARS-CoV polypeptide or a fragment, variant, or derivative thereof to a human, which comprise administering to a human one or more of the polynucleotide compositions described herein such that upon administration of polynucleotide compositions such as those described herein, a SARS-CoV polypeptide or a fragment, variant, or derivative thereof is expressed in human cells, in an amount sufficient to generate an immune response to SARS-CoV; or administering the SARS-CoV polypeptide or a fragment, variant, or derivative thereof itself to the human in an amount sufficient to generate an immune response.

[0236] The present invention further provides methods for delivering a SARS-CoV polypeptide or a fragment, variant, or derivative thereof to a human, which comprise administering to a vertebrate one or more of the compositions described herein; such that upon administration of compositions such as those described herein, an immune response is generated in the vertebrate.

[0237] The term "vertebrate" is intended to encompass a singular "vertebrate" as well as plural "vertebrates" and comprises in mammals and birds, as well as fish, reptiles, and amphibians.

[0238] The term "mammal" is intended to encompass a singular "mammal" and plural "mammals," and includes, but is not limited to humans; primates such as apes, monkeys (e.g., owl, squirrel, cebus, rhesus, African green, patas, cynomolgus, and cercopithecus), orangutans, baboons, gibbons, and chimpanzees; canids such as dogs and wolves; felids such as cats, lions, and tigers; equines such as horses, donkeys, and zebras, food animals such as cows, pigs, and sheep; ungulates such as deer and giraffes; ursids such as bears; and others such as rabbits, mice, ferrets, seals, whales. In particular, the mammal can be a human subject, a food animal or a companion animal.

[0239] The term "bird" is intended to encompass a singular "bird" and plural "birds," and includes, but is not limited to feral water birds such as ducks, geese, terns, shearwaters, and gulls; as well as domestic avian species such as turkeys, chickens, quail, pheasants, geese, and ducks. The term "bird" also encompasses passerine birds such as starlings and budgerigars.

[0240] The present invention further provides a method for generating, enhancing or modulating an immune response to SARS-CoV comprising administering to a vertebrate one or more of the compositions described herein. In this method, the compositions may include one or more isolated polynucleotides comprising at least one nucleic acid fragment where the nucleic acid fragment is a fragment of a coding region or a codon-optimized coding region encoding an SARS-CoV polypeptide, or a fragment, variant, or

derivative thereof. In another embodiment, the compositions may include multiple (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10) polynucleotides as described herein, such polynucleotides encoding different SARS CoV polypeptides in the same composition.

[0241] In another embodiment, the compositions may include both a polynucleotide as described above; and also an isolated SARS-CoV polypeptide, or a fragment, variant, or derivative thereof, wherein the protein is provided as a recombinant protein, in particular, a fusion protein, a purified subunit, viral vector expressing the protein, or inactivated virus. Thus, the latter compositions include both a polynucleotide encoding a SARS-CoV polypeptide or a fragment, variant, or derivative thereof and an isolated SARS-CoV polypeptide or a fragment, variant, or derivative thereof. The SARS-CoV polypeptide or a fragment, variant, or derivative thereof encoded by the polynucleotide of the compositions need not be the same as the isolated SARS-CoV polypeptide or a fragment, variant, or derivative thereof of the compositions. Compositions to be used according to this method may be univalent, bivalent, trivalent or multivalent.

[0242] The polynucleotides of the compositions may comprise a fragment of a coding region or a human (or other vertebrate) codon-optimized coding region encoding a protein of SARS-CoV, or a fragment, variant, or derivative thereof. The polynucleotides are incorporated into the cells of the vertebrate in vivo, and an antigenic amount of the SARS-CoV polypeptide, or fragment, variant, or derivative thereof, is produced in vivo. Upon administration of the composition according to this method, the SARS-CoV polypeptide or a fragment, variant, or derivative thereof is expressed in the vertebrate in an amount sufficient to elicit an immune response. Such an immune response might be used, for example, to generate antibodies to the SARS-CoV for use in diagnostic assays or as laboratory reagents, or as therapeutic or preventative vaccines as described herein.

[0243] The present invention further provides a method for generating, enhancing, or modulating a protective and/or therapeutic immune response to SARS-CoV in a vertebrate, comprising administering to a vertebrate in need of therapeutic and/or preventative immunity one or more of the compositions described herein. In this method, the compositions include one or more polynucleotides comprising at least one nucleic acid fragment, where the nucleic acid fragment is a fragment of a wild-type coding region or a codon-optimized coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof. In a further embodiment, the composition used in this method includes both an isolated polynucleotide comprising at least one nucleic acid fragment, where the nucleic acid fragment is a fragment of a wild-type coding region or a codon-optimized coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof; and at least one isolated SARS-CoV polypeptide, or a fragment, variant, or derivative thereof. Thus, the latter composition includes both an isolated polynucleotide encoding a SARS-CoV polypeptide or a fragment, variant, or derivative thereof and an isolated SARS-CoV polypeptide or a fragment, variant, or derivative thereof, for example, a recombinant protein, a purified subunit, or viral vector expressing the protein. Upon administration of the composition according to this method, the SARS-CoV polypeptide or a fragment, variant, or

derivative thereof is expressed in the vertebrate in a therapeutically or prophylactically effective amount.

[0244] In certain embodiments, the polynucleotide or polypeptide compositions of the present invention may be administered to a vertebrate where the vertebrate is used as an in vivo model to observe the effects of individual or multiple SARS-CoV polypeptides in vivo. This approach would not only eliminate the species specific barrier to studying SARS-CoV, but would allow for the study of the immunopathology of SARS-CoV polypeptides as well as SARS-CoV polypeptide specific effects without using infectious SARS-CoV virus. An in vivo vertebrate model of SARS infection would be useful, for example, in developing treatments for one or more aspects of SARS infection by mimicking those aspects of infection without the potential hazards associated with handling the infectious virus.

[0245] As used herein, an "immune response" refers to the ability of a vertebrate to elicit an immune reaction to a composition delivered to that vertebrate. Examples of immune responses include an antibody response or a cellular, e.g., T-cell, response. One or more compositions of the present invention may be used to prevent SARS-CoV infection in vertebrates, e.g., as a prophylactic or preventative vaccine (also sometimes referred to in the art as a "protective" vaccine), to establish or enhance immunity to SARS-CoV in a healthy individual prior to exposure to SARS-CoV or contraction of Severe Acute Respiratory Syndrome (SARS), thus preventing the syndrome or reducing the severity of SARS symptoms. As used herein, "a detectable immune response" refers to an immunogenic response to the polynucleotides and polypeptides of the present invention, which can be measured or observed by standard protocols. These protocols include, but are not limited to, immunoblot analysis (western), fluorescence-activated cell sorting (FACS), immunoprecipitation analysis, ELISA, cytolytic T-cell response, ELISPOT, and chromium release assay. An immune response may also be "detected" through challenge of immunized animals with virulent SARS-CoV, either before or after vaccination. ELISA assays are performed as described by Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1989). Cytolytic T-cell responses are measured as described in Hartikka et al., "Vaxfectin Enhances the Humoral Response to Plasmid DNA-encoded Antigens." *Vaccine* 19: 1911-1923 (2001), which is hereby incorporated in its entirety by reference. Standard ELISPOT technology is used for the CD4+ and CD8+ T-cell assays as described in Example 6A. Standard chromium release assays are used to measure specific cytotoxic T lymphocyte (CTL) activity against the various SARS-CoV antigens.

[0246] As mentioned above, compositions of the present invention may be used both to prevent SARS-CoV infection, and also to therapeutically treat SARS-CoV infection. In individuals already exposed to SARS-CoV, or already suffering from SARS, the present invention is used to further stimulate the immune system of the vertebrate, thus reducing or eliminating the symptoms associated with that disease or disorder. As defined herein, "treatment" refers to the use of one or more compositions of the present invention to prevent, cure, retard, or reduce the severity of SARS symptoms in a vertebrate, and/or result in no worsening of SARS over a specified period of time in a vertebrate which has already been exposed to SARS-CoV and is thus in need of

therapy. The term "prevention" refers to the use of one or more compositions of the present invention to generate immunity in a vertebrate which has not yet been exposed to a particular strain of SARS-CoV, thereby preventing or reducing disease symptoms if the vertebrate is later exposed to the particular strain of SARS-CoV. The methods of the present invention therefore may be referred to as therapeutic vaccination or preventative or prophylactic vaccination. It is not required that any composition of the present invention provide total immunity to SARS-CoV or totally cure or eliminate all SARS symptoms. As used herein, a "vertebrate in need of therapeutic and/or preventative immunity" refers to an individual for whom it is desirable to treat, i.e., to prevent, cure, retard, or reduce the severity of SARS symptoms, and/or result in no worsening of SARS over a specified period of time. Vertebrates to treat and/or vaccinate include humans, apes, monkeys (e.g., owl, squirrel, cebus, rhesus, African green, patas, cynomolgus, and cercopithecus), orangutans, baboons, gibbons, and chimpanzees, dogs, wolves, cats, lions, and tigers, horses, donkeys, zebras, cows, pigs, sheep, deer, giraffes, bears, rabbits, mice, ferrets, seals, whales, ducks, geese, terns, shearwaters, gulls, turkeys, chickens, quail, pheasants, geese, starlings and budgerigars.

[0247] One or more compositions of the present invention are utilized in a "prime boost" regimen. An example of a "prime boost" regimen may be found in Yang, Z. et al. *J. Virol.* 77:799-803 (2002). In these embodiments, one or more polynucleotide vaccine compositions of the present invention are delivered to a vertebrate, thereby priming the immune response of the vertebrate to SARS-CoV, and then a second immunogenic composition is utilized as a boost vaccination. One or more compositions of the present invention are used to prime immunity, and then a second immunogenic composition, e.g., a recombinant viral vaccine or vaccines, a different polynucleotide vaccine, or one or more purified subunit isolated SARS-CoV polypeptides or fragments, variants or derivatives thereof is used to boost the anti-SARS-CoV immune response.

[0248] In one embodiment, a priming composition and a boosting composition are delivered to a vertebrate in separate doses and vaccinations. For example, a single composition may comprise one or more polynucleotides encoding SARS-CoV protein(s), fragment(s), variant(s), or derivative(s) thereof and/or one or more isolated SARS-CoV polypeptide(s) or fragment(s), variant(s), or derivative(s) thereof as the priming component. The polynucleotides encoding the SARS-CoV polypeptides fragments, variants, or derivatives thereof may be contained in a single plasmid or viral vector or in multiple plasmids or viral vectors. At least one polynucleotide encoding a SARS-CoV protein and/or one or more SARS-CoV isolated polypeptide can serve as the boosting component. In this embodiment, the compositions of the priming component and the compositions of the boosting component may be contained in separate vials. In one example, the boosting component is administered approximately 1 to 6 months after administration of the priming component.

[0249] In one embodiment, a priming composition and a boosting composition are combined in a single composition or single formulation. For example, a single composition may comprise an isolated SARS-CoV polypeptide or a fragment, variant, or derivative thereof as the priming com-

ponent and a polynucleotide encoding an SARS-CoV protein as the boosting component. In this embodiment, the compositions may be contained in a single vial where the priming component and boosting component are mixed together. In general, because the peak levels of expression of protein from the polynucleotide does not occur until later (e.g., 7-10 days) after administration, the polynucleotide component may provide a boost to the isolated protein component. Compositions comprising both a priming component and a boosting component are referred to herein as "combinatorial vaccine compositions" or "single formulation heterologous prime-boost vaccine compositions." In addition, the priming composition may be administered before the boosting composition, or even after the boosting composition, if the boosting composition is expected to take longer to act.

[0250] In another embodiment, the priming composition may be administered simultaneously with the boosting composition, but in separate formulations where the priming component and the boosting component are separated.

[0251] The terms "priming" or "primary" and "boost" or "boosting" as used herein may refer to the initial and subsequent immunizations, respectively, i.e., in accordance with the definitions these terms normally have in immunology. However, in certain embodiments, e.g., where the priming component and boosting component are in a single formulation, initial and subsequent immunizations may not be necessary as both the "prime" and the "boost" compositions are administered simultaneously.

[0252] In certain embodiments, one or more compositions of the present invention are delivered to a vertebrate by methods described herein, thereby achieving an effective therapeutic and/or an effective preventative immune response. More specifically, the compositions of the present invention may be administered to any tissue of a vertebrate, including, but not limited to, muscle, skin, brain tissue, lung tissue, liver tissue, spleen tissue, bone marrow tissue, thymus tissue, heart tissue, e.g., myocardium, endocardium, and pericardium, lymph tissue, blood tissue, bone tissue, pancreas tissue, kidney tissue, gall bladder tissue, stomach tissue, intestinal tissue, testicular tissue, ovarian tissue, uterine tissue, vaginal tissue, rectal tissue, nervous system tissue, eye tissue, glandular tissue, tongue tissue, and connective tissue, e.g., cartilage.

[0253] Furthermore, the compositions of the present invention may be administered to any internal cavity of a vertebrate, including, but not limited to, the lungs, the mouth, the nasal cavity, the stomach, the peritoneal cavity, the intestine, any heart chamber, veins, arteries, capillaries, lymphatic cavities, the uterine cavity, the vaginal cavity, the rectal cavity, joint cavities, ventricles in brain, spinal canal in spinal cord, the ocular cavities, the lumen of a duct of a salivary gland or a liver. When the compositions of the present invention are administered to the lumen of a duct of a salivary gland or liver, the desired polypeptide is expressed in the salivary gland and the liver such that the polypeptide is delivered into the blood stream of the vertebrate from each of the salivary gland or the liver. Certain modes for administration to secretory organs of a gastrointestinal system using the salivary gland, liver and pancreas to release a desired polypeptide into the bloodstream are disclosed in U.S. Pat. Nos. 5,837,693 and 6,004,944, both of which are incorporated herein by reference in their entireties.

[0254] In certain embodiments, the compositions are administered to muscle, either skeletal muscle or cardiac muscle, or to lung tissue. Specific, but non-limiting modes for administration to lung tissue are disclosed in Wheeler, C. J., et al., *Proc. Natl. Acad. Sci. USA* 93:11454-11459 (1996), which is incorporated herein by reference in its entirety.

[0255] According to the disclosed methods, compositions of the present invention can be administered by intramuscular (i.m.), subcutaneous (s.c.), or intrapulmonary routes. Other suitable routes of administration include, but are not limited to intratracheal, transdermal, intracutaneous, intranasal, inhalation, intracavity, intravenous (i.v.), intraductal (e.g., into the pancreas) and intraparenchymal (i.e., into any tissue) administration. Transdermal delivery includes, but is not limited to intradermal (e.g., into the dermis or epidermis), transdermal (e.g., percutaneous) and transmucosal administration (i.e., into or through skin or mucosal tissue). Intracavity administration includes, but is not limited to administration into oral, vaginal, rectal, nasal, peritoneal, or intestinal cavities as well as, intrathecal (i.e., into spinal canal), intraventricular (i.e., into the brain ventricles or the heart ventricles), intraatrial (i.e., into the heart atrium) and sub arachnoid (i.e., into the sub arachnoid spaces of the brain) administration.

[0256] Any mode of administration can be used so long as the mode results in the expression of the desired peptide or protein, in the desired tissue, in an amount sufficient to generate an immune response to SARS-CoV and/or to generate a prophylactically or therapeutically effective immune response to SARS-CoV in a vertebrate in need of such response. Administration means of the present invention include needle injection, catheter infusion, biolistic injectors, particle accelerators (e.g., "gene guns" or pneumatic "needleless" injectors) Med-B-Jet (Vahlsing, H., et al., *J. Immunol. Methods* 171:11-22 (1994)), Pigjet (Schrijver, R., et al., *Vaccine* 15: 1908-1916 (1997)), Biojector (Davis, H., et al., *Vaccine* 12: 1503-1509 (1994)), Gramzinski, R., et al., *Mol. Med.* 4: 109-118 (1998)), AdvantaJet (Linnmayer, I., et al., *Diabetes Care* 9:294-297 (1986)), Medi-jector (Martins, J., and Reedl, E. *J. Occup. Med.* 21:821-824 (1979)), gel foam sponge deposits, other commercially available depot materials (e.g., hydrogels), osmotic pumps (e.g., Alza minipumps), oral or suppositories solid (tablet or pill) pharmaceutical formulations, topical skin creams, and decanting, use of polynucleotide coated surface (Qin, Y., et al., *Life Sciences* 65: 2193-2203 (1999)) or topical applications during surgery. Certain modes of administration are intramuscular needle-based injection and pulmonary application via catheter infusion. Energy-assisted plasmid delivery (EAPD) methods may also be employed to administer the compositions of the invention. One such method involves the application of brief electrical pulses to injected tissues, a procedure commonly known as electroporation. See generally Mir, L. M., et al., *Proc. Natl. Acad. Sci USA* 96:4262-7 (1999); Hartikka, J. et al., *Mol. Ther.* 4:407-15 (2001); Mathiesen, I., *Gene Ther.* 6:508-14 (1999); Rizzuto G. et al., *Hum. Gen. Ther.* 11:1891-900 (2000). Each of the references cited in this paragraph is incorporated herein by reference in its entirety.

[0257] Determining an effective amount of one or more compositions of the present invention depends upon a number of factors including, for example, the antigen being expressed or administered directly, (e.g., S, N, E or M, or

fragments, variants, or derivatives thereof), the age and weight of the subject, the precise condition requiring treatment and its severity, and the route of administration. Based on the above factors, determining the precise amount, number of doses, and timing of doses are within the ordinary skill in the art and will be readily determined by the attending physician or veterinarian.

[0258] Compositions of the present invention may include various salts, excipients, delivery vehicles and/or auxiliary agents as are disclosed, e.g., in U.S. Patent Application Publication 2002/0019358, published Feb. 14, 2002, which is incorporated herein by reference in its entirety.

[0259] Furthermore, compositions of the present invention may include one or more transfection facilitating compounds that facilitate delivery of polynucleotides to the interior of a cell, and/or to a desired location within a cell. As used herein, the terms "transfection facilitating compound," "transfection facilitating agent," and "transfection facilitating material" are synonymous, and may be used interchangeably. It should be noted that certain transfection facilitating compounds may also be "adjuvants" as described infra, i.e., in addition to facilitating delivery of polynucleotides to the interior of a cell, the compound acts to alter or increase the immune response to the antigen encoded by that polynucleotide. Examples of the transfection facilitating compounds include, but are not limited to inorganic materials such as calcium phosphate, alum (aluminum sulfate), and gold particles (e.g., "powder" type delivery vehicles); peptides that are, for example, cationic, intercell targeting (for selective delivery to certain cell types), intracellular targeting (for nuclear localization or endosomal escape), and amphoteric (helix forming or pore forming); proteins that are, for example, basic (e.g., positively charged) such as histones, targeting (e.g., asialoglycoprotein), viral (e.g., Sendai virus coat protein), and pore-forming; lipids that are, for example, cationic (e.g., DMRIE, DOSPA, DC-Chol), basic (e.g., sterylamine), neutral (e.g., cholesterol), anionic (e.g., phosphatidyl serine), and zwitterionic (e.g., DOPE, DOPC); and polymers such as dendrimers, star-polymers, "homogeneous" poly-amino acids (e.g., poly-lysine, poly-arginine), "heterogeneous" poly-amino acids (e.g., mixtures of lysine & glycine), co-polymers, polyvinylpyrrolidone (PVP), poloxamers (e.g., CRL 1005) and polyethylene glycol (PEG). A transfection facilitating material can be used alone or in combination with one or more other transfection facilitating materials. Two or more transfection facilitating materials can be combined by chemical bonding (e.g., covalent and ionic such as in lipidated polylysine, PEGylated polylysine) (Toucheva, et al., *Biochim. Biophys. Acta* 1380 (3):354-368 (1998)), mechanical mixing (e.g., free moving materials in liquid or solid phase such as "polylysine-cationic lipids") (Gao and Huang, *Biochemistry* 35:1027-1036 (1996); Trubetskoy, et al., *Biochem. Biophys. Acta* 1131:311-313 (1992)), and aggregation (e.g., co-precipitation, gel forming such as in cationic lipids+poly-lactide, and polylysine+gelatin).

[0260] One category of transfection facilitating materials is cationic lipids. Examples of cationic lipids are 5-carboxypermethylglycine dioctadecylamide (DOGS) and dipalmitoyl-phosphatidylethanolamine-5-carboxypermethylglycine (DPPES). Cationic cholesterol derivatives are also useful, including {3-[N,N'-dimethylamino]ethane}-carboxyl-cholesterol (DC-Chol). Dimethyl dioctadecyl-ammo-

nium bromide (DDAB), N-(3-aminopropyl)-N,N-bis-(2-tetradecyloxyethyl)-N-methyl-ammonium bromide (PA-DEMO), N-(3-aminopropyl)-N,N-bis-(2-dodecyloxyethyl)-N-methyl-ammonium bromide (PA-DELO), N,N,N-tris-(2-dodecyloxy)ethyl-N-(3-amino)propyl-ammonium bromide (PA-TELO), and N1-(3-aminopropyl)-(2-dodecyloxy)ethyl-N2-(2-dodecyloxy)ethyl-1-piperazinanium bromide (GA-LOE-BP) can also be employed in the present invention.

[0261] Non-diether cationic lipids, such as DL-1,2-dioleoyl-3-dimethylaminopropyl- β -hydroxyethylammonium (DORI diester), 1-O-oleyl-2-oleyl-3-dimethylaminopropyl- β -hydroxyethylammonium (DORI ester/ether), and their salts promote *in vivo* gene delivery. In some embodiments, cationic lipids comprise groups attached via a heteroatom attached to the quaternary ammonium moiety in the head group. A glyceryl spacer can connect the linker to the hydroxyl group.

[0262] Specific, but non-limiting cationic lipids for use in certain embodiments of the present invention include DMRIE ((α)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide), GAP-DMORIE ((α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecyloxy)-1-propanaminium bromide), and GAP-DLRIE ((α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide).

[0263] Other specific but non-limiting cationic surfactants for use in certain embodiments of the present invention include Bn-DHRIE, DlxRIE, DlxRIE-OAc, DlxRIE-Obz and Pr-DOctRIE-OAc. These lipids are disclosed in copending U.S. patent application No. {Attorney Docket No. 1530.0610000}. In another aspect of the present invention, the cationic surfactant is Pr-DOctRIE-OAc.

[0264] Other cationic lipids include (α)-N,N-dimethyl-N-[2-(sperminecarboxamido) ethyl]-2,3-bis(dioleoyl)-1-propanaminium pentahydrochloride (DOSPA), (α)-N-(2-aminoethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide (β -aminoethyl-DMRIE or β AE-DMRIE) (Wheeler, et al., *Biochim. Biophys. Acta* 1280:1-11 (1996)), and (α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide (GAP-DLRIE) (Wheeler, et al., *Proc. Natl. Acad. Sci. USA* 93:11454-11459 (1996)), which have been developed from DMRIE.

[0265] Other examples of DMRIE-derived cationic lipids that are useful for the present invention are (α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide (GAP-DDRIE), (α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide (GAP-DMRIE), (α)-N-(N^m-methyl)-N^m-ureylpropyl-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide (GMU-DMRIE), (α)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide (DLRIE), and (α)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis-[Z]-9-octadecyloxypropyl-1-propanaminium bromide (HP-DORIE).

[0266] In the embodiments where the immunogenic composition comprises a cationic lipid, the cationic lipid may be mixed with one or more co-lipids. For purposes of definition, the term "co-lipid" refers to any hydrophobic material which may be combined with the cationic lipid component and includes amphipathic lipids, such as phospholipids, and

neutral lipids, such as cholesterol. Cationic lipids and co-lipids may be mixed or combined in a number of ways to produce a variety of non-covalently bonded macroscopic structures, including, for example, liposomes, multilamellar vesicles, unilamellar vesicles, micelles, and simple films. One non-limiting class of co-lipids are the zwitterionic phospholipids, which include the phosphatidylethanolamines and the phosphatidylcholines. Examples of phosphatidylethanolamines, include DOPE, DMPE and DPPE. In certain embodiments, the co-lipid is DPPE, which comprises two phytanoyl substituents incorporated into the diacylphosphatidylethanolamine skeleton.

[0267] In other embodiments, the co-lipid is DOPE, CAS name 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine.

[0268] When a composition of the present invention comprises a cationic lipid and co-lipid, the cationic lipid:co-lipid molar ratio may be from about 9:1 to about 1:9, from about 4:1 to about 1:4, from about 2:1 to about 1:2, or about 1:1.

[0269] In order to maximize homogeneity, the cationic lipid and co-lipid components may be dissolved in a solvent such as chloroform, followed by evaporation of the cationic lipid:co-lipid solution under vacuum to dryness as a film on the inner surface of a glass vessel (e.g., a Rotavap round-bottomed flask). Upon suspension in an aqueous solvent, the amphipathic lipid component molecules self-assemble into homogeneous lipid vesicles. These lipid vesicles may subsequently be processed to have a selected mean diameter of uniform size prior to complexing with, for example, a polynucleotide or a codon-optimized polynucleotide of the present invention, according to methods known to those skilled in the art. For example, the evaporation of a lipid solution is described in Felgner et al., *Proc. Natl. Acad. Sci. USA* 8:7413-7417 (1987) and in U.S. Pat. No. 5,264,618, the disclosures of which are incorporated herein by reference.

[0270] In those embodiments where the composition includes a cationic lipid, polynucleotides of the present invention are complexed with lipids by mixing, for example, a plasmid in aqueous solution and a solution of cationic lipid:co-lipid as prepared herein are mixed. The concentration of each of the constituent solutions can be adjusted prior to mixing such that the desired final plasmid/cationic lipid:co-lipid ratio and the desired plasmid final concentration will be obtained upon mixing the two solutions. The cationic lipid:co-lipid mixtures are suitably prepared by hydrating a thin film of the mixed lipid materials in an appropriate volume of aqueous solvent by vortex mixing at ambient temperatures for about 1 minute. The thin films are prepared by admixing chloroform solutions of the individual components to afford a desired molar solute ratio followed by aliquoting the desired volume of the solutions into a suitable container. The solvent is removed by evaporation, first with a stream of dry, inert gas (e.g., argon) followed by high vacuum treatment.

[0271] Other hydrophobic and amphiphilic additives, such as, for example, sterols, fatty acids, gangliosides, glycolipids, lipopeptides, liposaccharides, neobees, niosomes, prostaglandins and sphingolipids, may also be included in compositions of the present invention. In such compositions, these additives may be included in an amount between about 0.1 mol % and about 99.9 mol % (relative to total lipid), about 1-50 mol %, or about 2-25 mol %.

[0272] Additional embodiments of the present invention are drawn to compositions comprising an auxiliary agent which is administered before, after, or concurrently with the polynucleotide. As used herein, an "auxiliary agent" is a substance included in a composition for its ability to enhance, relative to a composition which is identical except for the inclusion of the auxiliary agent, the entry of polynucleotides into vertebrate cells in vivo, and/or the in vivo expression of polypeptides encoded by such polynucleotides. Certain auxiliary agents may, in addition to enhancing entry of polynucleotides into cells, enhance an immune response to an immunogen encoded by the polynucleotide. Auxiliary agents of the present invention include nonionic, anionic, cationic, or zwitterionic surfactants or detergents, with nonionic surfactants or detergents being preferred, chelators, DNase inhibitors, poloxamers, agents that aggregate or condense nucleic acids, emulsifying or solubilizing agents, wetting agents, gel-forming agents, and buffers.

[0273] Auxiliary agents for use in compositions of the present invention include, but are not limited to non-ionic detergents and surfactants IGEPAL CA 630®, NONIDET NP-40, Nonidet® P40, Tween-20®, Tween-80™, Pluronic® F68 (ave. MW: 8400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 80%), Pluronic® F770® (ave. MW: 6600; approx. MW of hydrophobe, 2100; approx. wt. % of hydrophile, 70%), Pluronic® P65® (ave. MW: 3400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 50%), Triton X-100™, and Triton X-114™; the anionic detergent sodium dodecyl sulfate (SDS); the sugar stachyose; the condensing agent DMSO; and the chelator/DNase inhibitor EDTA, CRL 1005 (12 kDa, 5% POE), and BAK (Benzalkonium chloride 50% solution, available from Ruger Chemical Co. Inc.). In certain specific embodiments, the auxiliary agent is DMSO, Nonidet P40, Pluronic® P68® (ave. MW: 8400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 80%), Pluronic® F77® (ave. MW: 6600; approx. MW of hydrophobe, 2100; approx. wt. % of hydrophile, 70%), Pluronic® P65® (ave. MW: 3400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 50%), Pluronic® L64® (ave. MW: 2900; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 40%), and Pluronic® F108® (ave. MW: 14600; approx. MW of hydrophobe, 3000; approx. wt. % of hydrophile, 80%). See, e.g., U.S. Patent Application Publication No. 2002/0019358, published Feb. 14, 2002, which is incorporated herein by reference in its entirety.

[0274] Certain compositions of the present invention may further include one or more adjuvants before, after, or concurrently with the polynucleotide. The term "adjuvant" refers to any material having the ability to (1) alter or increase the immune response to a particular antigen or (2) increase or aid an effect of a pharmacological agent. It should be noted, with respect to polynucleotide vaccines, that an "adjuvant," may be a transfection facilitating material. Similarly, certain "transfection facilitating materials" described supra, may also be an "adjuvant." An adjuvant may be used with a composition comprising a polynucleotide of the present invention. In a prime-boost regimen, as described herein, an adjuvant may be used with either the priming immunization, the booster immunization, or both. Suitable adjuvants include, but are not limited to, cytokines and growth factors, bacterial components (e.g., endotoxins, in particular superantigens, exotoxins and cell wall components); aluminum-based salts; calcium-based salts; silica;

polynucleotides; toxoids; serum proteins, viruses and virally-derived materials, poisons, venoms, imidazoquinoline compounds, poloxamers, and cationic lipids.

[0275] A great variety of materials have been shown to have adjuvant activity through a variety of mechanisms. Any compound which may increase the expression, antigenicity or immunogenicity of the polypeptide is a potential adjuvant. The present invention provides an assay to screen for improved immune responses to potential adjuvants. Potential adjuvants which may be screened for their ability to enhance the immune response according to the present invention include, but are not limited to: inert carriers, such as alum, bentonite, latex, and acrylic particles; pluronic block polymers, such as TiterMax® (block copolymer CRL-8941, squalene (a metabolizable oil) and a microparticulate silica stabilizer), depot formers, such as Freunds adjuvant, surface active materials, such as saponin, lysolecithin, retinal, Quil A, liposomes, and pluronic polymer formulations; macrophage stimulators, such as bacterial lipopolysaccharide; alternate pathway complement activators, such as insulin, zymosan, endotoxin, and levamisole; and non-ionic surfactants, such as poloxamers, poly(oxyethylene)-poly(oxypropylene) tri-block copolymers. Also included as adjuvants are transfection-facilitating materials, such as those described above.

[0276] Poloxamers which may be screened for their ability to enhance the immune response according to the present invention include, but are not limited to, commercially available poloxamers such as Pluronic® surfactants, which are block copolymers of propylene oxide and ethylene oxide in which the propylene oxide block is sandwiched between two ethylene oxide blocks. Examples of Pluronic® surfactants include Pluronic® L121 (ave. MW: 4400; approx. MW of hydrophobe, 3600; approx. wt. % of hydrophile, 10%), Pluronic® L101 (ave. MW: 3800; approx. MW of hydrophobe, 3000; approx. wt. % of hydrophile, 10%), Pluronic® L81 (ave. MW: 2750; approx. MW of hydrophobe, 2400; approx. wt. % of hydrophile, 10%), Pluronic® L61 (ave. MW: 2000; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 10%), Pluronic® L31 (ave. MW: 1100; approx. MW of hydrophobe, 900; approx. wt. % of hydrophile, 10%), Pluronic® L122 (ave. MW: 5000; approx. MW of hydrophobe, 3600; approx. wt. % of hydrophile, 20%), Pluronic® L92 (ave. MW: 3650; approx. MW of hydrophobe, 2700; approx. wt. % of hydrophile, 20%), Pluronic® L72 (ave. MW: 2750; approx. MW of hydrophobe, 2100; approx. wt. % of hydrophile, 20%), Pluronic® L62 (ave. MW: 2500; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 20%), Pluronic® L42 (ave. MW: 1630; approx. MW of hydrophobe, 1200; approx. wt. % of hydrophile, 20%), Pluronic® L63 (ave. MW: 2650; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 30%), Pluronic® L43 (ave. MW: 1850; approx. MW of hydrophobe, 1200; approx. wt. % of hydrophile, 30%), Pluronic® L64 (ave. MW: 2900; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 40%), Pluronic® L44 (ave. MW: 2200; approx. MW of hydrophobe, 1200; approx. wt. % of hydrophile, 40%), Pluronic® L35 (ave. MW: 1900; approx. MW of hydrophobe, 900; approx. wt. % of hydrophile, 50%), Pluronic® P123 (ave. MW: 5750; approx. MW of hydrophobe, 3600; approx. wt. % of hydrophile, 30%), Pluronic® P103 (ave. MW: 4950; approx. MW of hydrophobe, 3000; approx. wt. % of hydrophile, 30%), Pluronic® P104 (ave. MW: 5900; approx. MW of hydrophobe, 3000;

approx. wt. % of hydrophile, 40%), Pluronic® P84 (ave. MW: 4200; approx. MW of hydrophobe, 2400; approx. wt. % of hydrophile, 40%), Pluronic® P105 (ave. MW: 6500; approx. MW of hydrophobe, 3000; approx. wt. % of hydrophile, 50%), Pluronic® P85 (ave. MW: 4600; approx. MW of hydrophobe, 2400; approx. wt. % of hydrophile, 50%), Pluronic® P75 (ave. MW: 4150; approx. MW of hydrophobe, 2100; approx. wt. % of hydrophile, 50%), Pluronic® P65 (ave. MW: 3400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 50%), Pluronic® F127 (ave. MW: 12600; approx. MW of hydrophobe, 3600; approx. wt. % of hydrophile, 70%), Pluronic® F98 (ave. MW: 13000; approx. MW of hydrophobe, 2700; approx. wt. % of hydrophile, 80%), Pluronic® F87 (ave. MW: 7700; approx. MW of hydrophobe, 2400; approx. wt. % of hydrophile, 70%), Pluronic® F77 (ave. MW: 6600; approx. MW of hydrophobe, 2100; approx. wt. % of hydrophile, 70%), Pluronic® F108 (ave. MW: 14600; approx. MW of hydrophobe, 3000; approx. wt. % of hydrophile, 80%), Pluronic® F98 (ave. MW: 13000; approx. MW of hydrophobe, 2700; approx. wt. % of hydrophile, 80%), Pluronic® F88 (ave. MW: 11400; approx. MW of hydrophobe, 2400; approx. wt. % of hydrophile, 80%), Pluronic® F68 (ave. MW: 8400; approx. MW of hydrophobe, 1800; approx. wt. % of hydrophile, 80%), Pluronic® F38 (ave. MW: 4700; approx. MW of hydrophobe, 900; approx. wt. % of hydrophile, 80%).

[0277] Reverse poloxamers which may be screened for their ability to enhance the immune response according to the present invention include, but are not limited to Pluronic® R 31R1 (ave. MW: 3250; approx. MW of hydrophobe, 3100; approx. wt. % of hydrophile, 10%), Pluronic® R 25R1 (ave. MW: 2700; approx. MW of hydrophobe, 2500; approx. wt. % of hydrophile, 10%), Pluronic® R 17R1 (ave. MW: 1900; approx. MW of hydrophobe, 1700; approx. wt. % of hydrophile, 10%), Pluronic® R 31R2 (ave. MW: 3300; approx. MW of hydrophobe, 3100; approx. wt. % of hydrophile, 20%), Pluronic® R 25R2 (ave. MW: 3100; approx. MW of hydrophobe, 2500; approx. wt. % of hydrophile, 20%), Pluronic® R 17R2 (ave. MW: 2150; approx. MW of hydrophobe, 1700; approx. wt. % of hydrophile, 20%), Pluronic® R 12R3 (ave. MW: 1800; approx. MW of hydrophobe, 1200; approx. wt. % of hydrophile, 30%), Pluronic® R 31R4 (ave. MW: 4150; approx. MW of hydrophobe, 3100; approx. wt. % of hydrophile, 40%), Pluronic® R 25R4 (ave. MW: 3600; approx. MW of hydrophobe, 2500; approx. wt. % of hydrophile, 40%), Pluronic® R 22R4 (ave. MW: 3350; approx. MW of hydrophobe, 2200; approx. wt. % of hydrophile, 40%), Pluronic® R 17R4 (ave. MW: 3650; approx. MW of hydrophobe, 1700; approx. wt. % of hydrophile, 40%), Pluronic® R 25R5 (ave. MW: 4320; approx. MW of hydrophobe, 2500; approx. wt. % of hydrophile, 50%), Pluronic® R 10R5 (ave. MW: 1950; approx. MW of hydrophobe, 1000; approx. wt. % of hydrophile, 50%), Pluronic® R 25R8 (ave. MW: 8550; approx. MW of hydrophobe, 2500; approx. wt. % of hydrophile, 80%), Pluronic® R 17R8 (ave. MW: 7000; approx. MW of hydrophobe, 1700; approx. wt. % of hydrophile, 80%), and Pluronic® R 10R8 (ave. MW: 4550; approx. MW of hydrophobe, 1000; approx. wt. % of hydrophile, 80%).

[0278] Other commercially available poloxamers which may be screened for their ability to enhance the immune response according to the present invention include compounds that are block copolymer of polyethylene and polypropylene glycol such as Syneronic® L121 (ave. MW:

4400, Syneronic® L122 (ave. MW: 5000, Syneronic® P104 (ave. MW: 5850, Syneronic® P105 (ave. MW: 6500), Syneronic® P123 (ave. MW: 5750), Syneronic® P85 (ave. MW: 4600) and Syneronic® P94 (ave. MW: 4600), in which L indicates that the surfactants are liquids, P that they are pastes, the first digit is a measure of the molecular weight of the polypropylene portion of the surfactant and the last digit of the number, multiplied by 10, gives the percent ethylene oxide content of the surfactant; and compounds that are nonylphenyl polyethylene glycol such as Syneronic® NP10 (nonylphenyl ethoxylated surfactant—10% solution), Syneronic® NP30 (condensate of 1 mole of nonylphenol with 30 moles of ethylene oxide) and Syneronic® NP5 (condensate of 1 mole of nonylphenol with 5.5 moles of naphthalene oxide).

[0279] Other poloxamers which may be screened for their ability to enhance the immune response according to the present invention include: (a) a polyether block copolymer comprising an A-type segment and a B-type segment, wherein the A-type segment comprises a linear polymeric segment of relatively hydrophilic character, the repeating units of which contribute an average Henschel-Les fragmental constant of about -0.4 or less and have molecular weight contributions between about 30 and about 500, wherein the B-type segment comprises a linear polymeric segment of relatively hydrophobic character, the repeating units of which contribute an average Henschel-Les fragmental constant of about -0.4 or more and have molecular weight contributions between about 30 and about 500, wherein at least about 80% of the linkages joining the repeating units for each of the polymeric segments comprise an ether linkage; (b) a block copolymer having a polyether segment and a polycation segment, wherein the polyether segment comprises at least an A-type block, and the polycation segment comprises a plurality of cationic repeating units; and (c) a polyether-polycation copolymer comprising a polymer, a polyether segment and a polycationic segment comprising a plurality of cationic repeating units of formula —NH—R^n , wherein R^n is a straight chain aliphatic group of 2 to 6 carbon atoms, which may be substituted, wherein said polyether segments comprise at least one of an A-type of B-type segment. See U.S. Pat. No. 5,656,611, by Kabonov, et al., which is incorporated herein by reference in its entirety. Other poloxamers of interest include CRL1005 (12 kDa, 5% POE), CRL8300 (11 kDa, 5% POE), CRL2690 (12 kDa, 10% POE), CRL4505 (15 kDa, 5% POE) and CRL1415 (9 kDa, 10% POE).

[0280] Other auxiliary agents which may be screened for their ability to enhance the immune response according to the present invention include, but are not limited to Acacia (gum arabic); the polyoxyethylene ether $\text{R—O—(C}_2\text{H}_4\text{O)}_x\text{—H}$ (BRJ®), e.g., polyethylene glycol dodecyl ether (BRJ® 35, x=23), polyethylene glycol dodecyl ether (BRJ® 30, x=4), polyethylene glycol hexadecyl ether (BRJ® 52, x=2), polyethylene glycol hexadecyl ether (BRJ® 56, x=10), polyethylene glycol hexadecyl ether (BRJ® 58P, x=20), polyethylene glycol octadecyl ether (BRJ® 72, x=2), polyethylene glycol octadecyl ether (BRJ® 76, x=10), polyethylene glycol octadecyl ether (BRJ® 78P, x=20), polyethylene glycol oleyl ether (BRJ® 92V, x=2), and polyoxyl 10 oleyl ether (BRJ® 97, x=10); poly-D-glucosamine (chitosan); chlorbutanol; cholesterol; diethanolamine; digitonin; dimethylsulfoxide (DMSO); ethylenediamine tetracetic acid (EDTA); glyceryl monoster-

ate; lanolin alcohols; mono- and di-glycerides; monoethanolamine; nonylphenol polyoxyethylene ether (NP-40®); octylphenoxypolyoxyethanol (NONIDET NP-40 from Amresco); ethyl phenol poly (ethylene glycol ether)ⁿ, n=11 (Nonidet® P40 from Roche); octyl phenol ethylene oxide condensate with about 9 ethylene oxide units (nonidet P40); IGEPAL CA 630® (octyl phenoxy) polyethoxyethanol; structurally same as NONIDET NP-40; oleic acid; oleyl alcohol; polyethylene glycol 8000; polyoxyl 20 cetostearyl ether; polyoxyl 35 castor oil; polyoxyl 40 hydrogenated castor oil; polyoxyl 40 stearate; polyoxyethylene sorbitan monolaurate (poly sorbate 20, or TWEEN-20®; polyoxyethylene sorbitan monooleate (poly sorbate 80, or TWEEN-80®); propylene glycol diacetate; propylene glycol monostearate; protamine sulfate; proteolytic enzymes; sodium dodecyl sulfate (SDS); sodium monolaurate; sodium stearate; sorbitan derivatives (SPAN®), e.g., sorbitan monopalmitate (SPAN® 40), sorbitan monostearate (SPAN® 60), sorbitan tristearate (SPAN® 65), sorbitan monooleate (SPAN® 80), and sorbitan trioleate (SPAN® 85); 2,6,10,15,19,23-hexamethyl-2,6,10,14,18,22-tetracosahexaene (squalene); stachyose; stearic acid; sucrose; surfactin (lipopeptide antibiotic from *Bacillus subtilis*); dodecylpoly(ethylene glycol) ether (Thesic®) MW 582.9; octyl phenol ethylene oxide condensate with about 9-10 ethylene oxide units (Triton X-100™); octyl phenol ethylene oxide condensate with about 7-8 ethylene oxide units (Triton X-114™); tris(2-hydroxyethyl)amine (trolamine); and emulsifying wax.

[0281] In certain adjuvant compositions, the adjuvant is a cytokine. A composition of the present invention can comprise one or more cytokines, chemokines, or compounds that induce the production of cytokines and chemokines, or a polynucleotide encoding one or more cytokines, chemokines, or compounds that induce the production of cytokines and chemokines. Examples include, but are not limited to granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), colony stimulating factor (CSF), erythropoietin (EPO), interleukin 2 (IL-2), interleukin-3 (IL-3), interleukin 4 (IL-4), interleukin 5 (IL-5), interleukin 6 (IL-6), interleukin 7 (IL-7), interleukin 8 (IL-8), interleukin 10 (IL-10), interleukin 12 (IL-12), interleukin 15 (IL-15), interleukin 18 (IL-18), interferon alpha (IFN α), interferon beta (IFN β), interferon gamma (IFN γ), interferon omega (IFN ω), interferon tau (IFN τ), interferon gamma inducing factor I (IGIF), transforming growth factor beta (TGF- β), RANTES (regulated upon activation, normal T-cell expressed and presumably secreted), macrophage inflammatory proteins (e.g., MIP-1 alpha and MIP-1 beta), *Leishmania* elongation initiating factor (LEIF), and Flt-3 ligand.

[0282] In certain compositions of the present invention, the polynucleotide construct may be complexed with an adjuvant composition comprising (x)-N-(3-aminopropyl)-N, N-dimethyl-2,3-bis(syn-9-tetradeceneyloxy)-1-propan-aminium bromide (GAP-DMORIE). The composition may also comprise one or more co-lipids, e.g., 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (DIPPE), and/or 1,2-dimyristoyl-glycer-3-phosphoethanolamine (DMPE). An adjuvant composition comprising: GAP-DMORIE and DIPPE at a 1:1 molar ratio is referred to herein as Vaxfectin™. See, e.g., PCT Publication No. WO 00/57917, which is incorporated herein by reference in its entirety.

tin™. See, e.g., PCT Publication No. WO 00/57917, which is incorporated herein by reference in its entirety.

[0283] In other embodiments, the polynucleotide itself may function as an adjuvant as is the case when the polynucleotides of the invention are derived, in whole or in part, from bacterial DNA. Bacterial DNA containing motifs of unmethylated CpG-dinucleotides (CpG-DNA) triggers innate immune cells in vertebrates through a pattern recognition receptor (including toll receptors such as TLR 9) and thus possesses potent immunostimulatory effects on macrophages, dendritic cells and B-lymphocytes. See, e.g., Wagner, H., *Curr. Opin. Microbiol.* 5:62-69 (2002); Jung, J. et al., *J. Immunol.* 169: 2368-73 (2002); see also Klinman, D. M. et al., *Proc. Natl. Acad. Sci. U.S.A.* 93:2879-83 (1996). Methods of using unmethylated CpG-dinucleotides as adjuvants are described in, for example, U.S. Pat. Nos. 6,207, 646, 6,406,705, and 6,429,199, the disclosures of which are herein incorporated by reference.

[0284] The ability of an adjuvant to increase the immune response to an antigen is typically manifested by a significant increase in immune-mediated protection. For example, an increase in humoral immunity is typically manifested by a significant increase in the titer of antibodies raised to the antigen, and an increase in T-cell activity is typically manifested in increased cell proliferation, or cellular cytotoxicity, or cytokine secretion. An adjuvant may also alter an immune response, for example, by changing a primarily humoral or Th₂ response into a primarily cellular, or Th₁ response.

[0285] In certain embodiments, the compositions of the present invention may be administered in the absence of one or more transfection facilitating materials or auxiliary agents. It has been shown that, surprisingly, the cells of living vertebrates are capable of taking up and expressing polynucleotides that have been injected in vivo, even in the absence of any agent to facilitate transfection. Cohen, J., *Science* 259: 1691-1692; Felgner, P., *Scientific American* 276: 102-106 (1997). These references are hereby incorporated by reference in their entireties. Thus, by way of non-limiting examples, nucleic acid molecules and/or polynucleotides of the present invention (e.g., plasmid DNA, mRNA, linear DNA, or oligonucleotides) may be administered in the absence of any one of, or any combination of more than one of the following transfection facilitating materials or auxiliary agents as described herein: inorganic materials including but not limited to calcium phosphate, alum, and/or gold particles; peptides including but not limited to cationic peptides, amphipathic peptides, intercell targeting peptides, and/or intracellular targeting peptides; proteins, including but not limited to basic (i.e., positively charged) proteins, targeting proteins, viral proteins, and/or pore-forming proteins; lipids, including but not limited to cationic lipids, anionic lipids, basic lipids, neutral lipids, and/or zwitterionic lipids; polymers including but not limited to dendrimers, star-polymers, "homogeneous" poly-amino acids, "heterogeneous" poly-amino acids, co-polymers, PVP, poloxamers, and/or PEG; surfactants, including but not limited to anionic surfactants, cationic surfactants, and zwitterionic surfactants; detergents, including but not limited to anionic detergents, cationic detergents, and zwitterionic detergents; chelators, including but not limited to EDTA; DNase inhibitors; condensing agents including, but not limited to DMSO; emulsifying or solubilizing agents; gel-forming agents; buffers, and/or adjuvants.

[0286] Nucleic acid molecules and/or polynucleotides of the present invention, e.g., plasmid DNA, mRNA, linear DNA or oligonucleotides, may be solubilized in any of various buffers. Suitable buffers include, for example, phosphate buffered saline (PBS), normal saline, Tris buffer, and sodium phosphate (e.g., 150 mM sodium phosphate). Insoluble polynucleotides may be solubilized in a weak acid or weak base, and then diluted to the desired volume with a buffer. The pH of the buffer may be adjusted as appropriate. In addition, a pharmaceutically acceptable additive can be used to provide an appropriate osmolality. Such additives are within the purview of one skilled in the art. For aqueous compositions used in vivo, sterile pyrogen-free water can be used. Such formulations will contain an effective amount of a polynucleotide together with a suitable amount of an aqueous solution in order to prepare pharmaceutically acceptable compositions suitable for administration to a human.

[0287] Compositions of the present invention can be formulated according to known methods. Suitable preparation methods are described, for example, in Remington's Pharmaceutical Sciences, 16th Edition, A. Osol, ed., Mack Publishing Co., Easton, Pa. (1980), and Remington's Pharmaceutical Sciences, 19th Edition, A. R. Gennaro, ed., Mack Publishing Co., Easton, Pa. (1995), both of which are incorporated herein by reference in their entireties. Although the composition may be administered as an aqueous solution, it can also be formulated as an emulsion, gel, solution, suspension, lyophilized form, or any other form known in the art. In addition, the composition may contain pharmaceutically acceptable additives including, for example, diluents, binders, stabilizers, and preservatives.

Passive Immunotherapy

[0288] Antibody therapy can be subdivided into two principally different activities: (i) passive immunotherapy using intact non-labeled antibodies or labeled antibodies and (ii) active immunotherapy using anti-idiotypes for re-establishment of network balance in autoimmunity

[0289] In passive immunotherapy, naked antibodies are administered to neutralize an antigen or to direct effector functions to targeted membrane associated antigens. Neutralization would be of a lymphokine, a hormone, or an anaphylatoxin, i.e., C5a. Effector functions include complement fixation, macrophage activation and recruitment, and antibody-dependent cell-mediated cytotoxicity (ADCC). Naked antibodies have been used to treat leukemia (Ritz, S.F. et al. *Blood*, 58:141-152 (1981)) and antibodies to GD2 have been used in treatments of neuroblastomas (Schultz et al. *Cancer Res.* 44:5914 (1984)) and melanomas (Irie et al., *Proc. Natl. Acad. Sci.* 83: 8694 (1986)). One major advantage of passive antibody immunization is that it provides immediate immunity that can last for weeks and possibly months. Casadevall, A. "Passive Antibody Administration (Immediate Immunity) as a Specific Defense against Biological Weapons." *Emerging Infectious Diseases*. 8:833-841 (2002).

[0290] The invention also provides for antibodies specifically reactive with SARS Co-V polypeptides which have been produced from an immune response elicited by the administration, to a vertebrate, of polynucleotide and polypeptides of the present invention. Anti-protein/antipeptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Labo-*

ratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A vertebrate such as a mouse, a hamster, a rabbit, a horse, a human, or non-human primate can be immunized with an immunogenic form of a SARS Co-V polypeptide or polynucleotide, of the present invention, encoding an immunogenic form of a SARS-CoV polypeptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the SARS-CoV polypeptide can be administered in the presence of adjuvant and as part of compositions described herein. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

[0291] The antibodies of the invention are immunospecific for antigenic determinants of the SARS-CoV polypeptides of the invention, e.g., antigenic determinants of a polypeptide of the invention or a closely related human or non-human mammalian homolog (e.g., 90% homologous and at least about 95% homologous). In an alternative embodiment of the invention, the SARS Co-V antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence of the invention. By "not substantially cross react," is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, less than 5 percent, or less than 1 percent, of the binding affinity for a protein of the invention. In an alternative embodiment, there is no cross-reactivity between viral and inammalian antigens.

[0292] In one embodiment, purified monoclonal antibodies or polyclonal antibodies containing the variable heavy and light sequences are used as therapeutic and prophylactic agents to treat or prevent SARS-CoV infection by passive antibody therapy. In general, this will comprise administering a therapeutically or prophylactically effective amount of the monoclonal or polyclonal antibodies to a susceptible vertebrate or one exhibiting SARS Co-V infection. A dosage effective amount will range from about 50 to 20,000 µg/Kg, and from about 100 to 5000 µg/Kg. However, suitable dosages will vary depending on factors such as the condition of the treated host, weight, etc. Suitable effective dosages may be determined by those skilled in the art.

[0293] In an alternative embodiment, purified antibodies and the polynucleotides or polypeptides of the present invention are administered simultaneously (at the same time) or subsequent to the administration of the isolated antibodies, thereby providing both immediate and long lasting protection.

[0294] The monoclonal or polyclonal antibodies may be administered by any mode of administration suitable for administering antibodies. Typically, the subject antibodies will be administered by injection, e.g., intravenous, intramuscular, or intraperitoneal injection (as described previously), or aerosol. Aerosol administration is particularly preferred if the subjects treated comprise newborn infants.

[0295] Formulation of antibodies in pharmaceutically acceptable form may be effected by known methods, using known pharmaceutical carriers and excipients. Suitable carriers and excipients include by way of non-limiting example buffered saline, and bovine serum albumin.

[0296] Any polynucleotides or polypeptides, as described herein, can be used to produce the isolated antibodies of the invention. For example, SARS-CoV proteins S, N, M, and E, fragments, variants and derivatives thereof, are purified as described in Example 2. The purified protein then serves as an antigen for producing SARS-CoV specific monoclonal and polyclonal antibodies.

[0297] Any vertebrate can serve as a host for antibody production. Preferred hosts include, but are not limited to human, non-human primate, mouse, rabbit, horse, goat, donkey, cow, sheep, chickens, cat, dog. Alternatively, antibodies can be produced by cultivation *ex vivo* of lymphocytes from primed donors stimulated with CD40 resulting in expansion of human B cells Banchereau et al., *Science* 251:70 (1991); Zhani et al., *J. Immunol.* 144:2955-2960, (1990); Tohma et al., *J. Immunol.* 146:2544-2552 (1991). Furthermore, an extra *in vitro* booster step can be used to obtain a higher yield of antibodies prior to immortalization of the cells. See Chaudhuri et al., *Cancer Supplement* 73: 1098-1104 (1994); Steenbakkers et al. *Hum. Antibod. Hybridomas* 4: 166-173 (1993); Ferraro et al., *Hum. Antibod. Hybridomas* 4:80-85 (1993); Kwekkeboom et al., *Immunol. Methods* 160:117-127 (1993), which are herein incorporated by reference.

[0298] An alternative to human primed donors, is to "recreate" or mimic splenic conditions in an immunocompromised animal host, such as the "Severe Combined Immune Deficient" (SCID) mouse. Human lymphocytes are readily adopted by the SCID mouse (hu-SCID) and produce high levels of immunoglobulins Mosier et al., *Nature* 335:256 (1988); McCune et al., *Science* 241:1632-1639 (1988). Moreover, if the donor used for reconstitution has been exposed to a particular antigen, a strong secondary response to the same antigen can be elicited in such mice. Duchosal et al. *Nature* 355:258-262 (1992).

[0299] The term "antibody" as used herein is intended to include fragments thereof which are also specifically reactive with SARS-CoV polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab'), fragments can be generated by treating antibody with pepsin. The resulting F(ab')₂ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-SARS-CoV portion.

[0300] Both monoclonal and polyclonal antibodies (Ab) directed against SARS-CoV polypeptides or SARS-CoV polypeptide variants, and antibody fragments such as Fab' and F(ab')₂, can be used to block the action of SARS-CoV polypeptides and allow the study of the role of a particular SARS-CoV polypeptide of the invention in the infectious life cycle of the virus and in pathogenesis.

[0301] Moreover, the antibodies possess utility as immunoprobe for diagnosis of SARS Co-V infection. This generally comprises taking a sample, e.g., respiratory fluid, of a person suspected of having SARS-CoV infection and incubating the sample with the subject human monoclonal antibodies to detect the presence of SARS-CoV infected cells. This involves directly or indirectly labeling the subject human antibodies with a reporter molecule which provides for detection of human monoclonal antibody SARS-CoV

immune complexes. Examples of known labels include by way of non-limiting example enzymes, e.g., β -lactamase, luciferase, and radiolabels. Methods for effecting immunodetection of antigens using monoclonal antibodies are well known in the art.

[0302] The following examples are included for purposes of illustration only and are not intended to limit the scope of the present invention, which is defined by the appended claims. All references cited in the Examples are incorporated herein by reference in their entireties.

EXAMPLES

Materials and Methods

[0303] The following materials and methods apply generally to all the examples disclosed herein. Specific materials and methods are disclosed in each example, as necessary.

[0304] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology (including PCR), vaccinology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., Sambrook et al., ed., Cold Spring Harbor Laboratory Press: (1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Millis et al. U.S. Pat. No. 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (JRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochromatography In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); and in Ausubel et al., *Current Protocols In Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1989).

Gene Construction

[0305] Constructs of the present invention are constructed based on the sequence information provided herein or in the art utilizing standard molecular biology techniques, including, but not limited to the following. First, a series complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of the construct are synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends. The single-stranded ends of each pair of oligonucleotides are designed to anneal with a single-stranded end of an adjacent oligonucleotide duplex. Several adjacent oligonucleotide pairs prepared in this manner are allowed to anneal, and approximately five to six adjacent oligonucleotide duplex fragments are then allowed to anneal together via the cohesive single stranded ends. This series of annealed oligonucleotide duplex fragments is then ligated together and cloned into a suitable plasmid, such as the TOPO® vector available from Invitrogen Corporation,

Carlsbad, Calif. The construct is then sequenced by standard methods. Constructs prepared in this manner, comprising 5 to 6 adjacent 80 to 90 base pair fragments ligated together, i.e., fragments of about 500 base pairs, are prepared, such that the entire desired sequence of the construct is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced. The oligonucleotides and primers referred to herein can easily be designed by a person of skill in the art based on the sequence information provided herein and in the art, and such can be synthesized by any of a number of commercial nucleotide providers, for example Retrogen, San Diego, Calif.

Plasmid Vector

[0306] Constructs of the present invention can be inserted, for example, into eukaryotic expression vectors VR1012 or VR1051. These vectors are built on a modified pUC18 background (see Yanisch-Perron, C., et al. *Gene* 33:103-119 (1985)), and contain a kanamycin resistance gene, the human cytomegalovirus immediate early promoter/enhancer and intron A, and the bovine growth hormone transcription termination signal, and a polylinker for inserting foreign genes. See Hartikka, J., et al., *Hum. Gene Ther.* 7:1205-1217 (1996). However, other standard commercially available eukaryotic expression vectors may be used in the present invention, including, but not limited to: plasmids pcDNA3, pCMV/Zeo, pCR3.1, pEF1/His, pIND/GS, pRC/HCMV2, pSV40/Zeo2, pTRACER-HCMV, pUB6/V5-His, pVAX1, and pZeoSV2 (available from Invitrogen, San Diego, Calif.), and plasmid pCI (available from Promega, Madison, Wis.).

[0307] An optimized backbone plasmid, termed VR-10551 has minor changes from the VR-1012 backbone described above. The VR-10551 vector is derived from and similar to VR-1012 in that it uses the human cytomegalovirus immediate early (hCMV-IE) gene enhancer/promoter and 5'untranslated region (UTR), including the hCMV-IE Intron A. The changes from the VR-1012 to the VR-10551 include some modifications to the multiple cloning site, and a modified rabbit globin 3'untranslated region/polyadenylation signal sequence/transcriptional terminator has been substituted for the same functional domain derived from the bovine growth hormone gene.

Plasmid DNA Purification

[0308] Plasmid DNA may be transformed into competent cells of an appropriate *Escherichia coli* strain (including but not limited to the DH5 α strain) and highly purified covalently closed circular plasmid DNA may be isolated by a modified lysis procedure (Horn, N. A., et al., *Hum. Gene Ther.* 6:565-573 (1995)) followed by standard double CsCl-ethidium bromide gradient ultracentrifugation (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989)). Alternatively, plasmid DNAs are purified using Giga columns from Qiagen (Valencia, Calif.) according to the kit instructions. All plasmid preparations are free of detectable chromosomal DNA, RNA and protein impurities based on gel analysis and the bicinchoninic protein assay (Pierce Chem. Co., Rockford Ill.). Endotoxin levels are measured using *Limulus* Amebocyte Lysate assay (LAL, Associates of Cape Cod, Falmouth, Mass.) in Endotoxin

Units/mg of plasmid DNA. The spectrophotometric A_{280}/A_{260} ratios of the DNA solutions are also determined. Plasmids are ethanol precipitated and resuspended in an appropriate solution, e.g., 150 mM sodium phosphate (for other appropriate excipients and auxiliary agents, see U.S. Patent Application Publication 20020019358, published Feb. 14, 2002). DNA is stored at -20°C until use. DNA is diluted by mixing it with 300 mM salt solutions and by adding appropriate amount of USP water to obtain 1 mg/ml plasmid DNA in the desired salt at the desired molar concentration.

Injections of Plasmid DNA

[0309] The quadriceps muscles of restrained awake mice (e.g., female 6-12 week old BALB/c mice from Harlan Sprague Dawley, Indianapolis, Ind.) are injected bilaterally with 50 μ g of DNA in 50 μ l solution (100 μ g in 100 μ l total per mouse) using a disposable plastic insulin syringe and 28G $\frac{1}{2}$ needle (Becton-Dickinson, Franklin Lakes, N.J., Cat. No. 329430) fitted with a plastic collar cut from a micropipette tip, as previously described (Hartikka, J., et al., *Hum. Gene Ther.* 7:1205-1217 (1996)).

[0310] Animal care will comply with the "Guide for the Use and Care of Laboratory Animals," Institute of Laboratory Animal Resources, Commission on Life Sciences, National Research Council, National Academy Press, Washington, D.C., 1996 as well as with Vical's Institutional Animal Care and Use Committee.

Example 1

Construction of Expression Vectors

[0311] Plasmid constructs comprising the native coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, are constructed as follows. The S, S1, S2, N, M, or E genes from SARS-CoV Urbani or other strains (e.g., CUHK-Su10, TOR2 and B01) are isolated from viral RNA by RT-PCR, or prepared by direct synthesis if the wildtype sequence is known, by standard methods and are inserted into the vector VR-10551 via standard restriction sites, by standard methods.

[0312] Plasmid constructs comprising human codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, are prepared as follows. The codon-optimized coding regions are generated using the full, minimal, uniform, or other codon optimization methods described herein. The coding regions or codon optimized coding regions are constructed using standard PCR methods described herein, or are ordered commercially. The coding regions or codon-optimized coding regions are inserted into the vector VR-10551 via standard restriction sites, by standard methods.

[0313] Examples of constructs to be made are listed in Table 19.

TABLE 19

Gene	Strain	Backbone	Wild type/Codon optimized
S	Urbani	10551	Wild type
S	Urbani	10551	Codon optimized
S1	Urbani	1012	Wild type
S1	Urbani	10551	Codon optimized
S2	Urbani	10551	Wild type
S2	Urbani	10551	Codon optimized
N	Urbani	10551	Wild type
N	Urbani	10551	Codon optimized
M	Urbani	10551	Wild type
M	Urbani	10551	Codon optimized
E	Urbani	10551	Wild type
E	Urbani	10551	Codon optimized

[0314] Plasmids constructed as above are propagated in *Escherichia coli* and purified by the alkaline lysis method (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., ed. 2 (1989)). CsCl-banded DNA are ethanol precipitated and resuspended in 0.9% saline to a final concentration of 2 mg/ml for injection. Alternately, plasmids are purified using any of a variety of commercial kits, or by other known procedures involving differential precipitation and/or chromatographic purification.

[0315] Expression is tested by formulating each of the plasmids in DMRIE/DOPE and transfecting cell lines including, but not limited to VM92 cells, fungal cells, including yeast cells such as *Saccharomyces* spp. cells; insect cells such as *Drosophila* S2, *Spodoptera* Sf9 or Sf21 cells and *Trichoplusia* High-Five cells; other animal cells (particularly mammalian cells and human cells) such as MDCK, CV1, 3T3, CPAE, A10, Sp2/0-Ag14, PC12, CHO, COS, VERO, HeLa, Bowes melanoma cells, SW-13, NCI-H295, RT4, HT-1376, UM-UC-3, IM-9, KG-1, R54;11, A-172, U-87MG, BT-20, MCF-7, SK-BR-3, ChaGo K-1, CCD-14Br, CaSki, ME-180, FHC, HT-29, Caco-2, SW480, HuTu80, Tera 1, NTERA-2, AN3 CA, KLE, RL95-2, Caki-1, ACHN, 769 P, CCRF-CEM, Hut 78, MOLT 4, HL-60, Hep-3B, HepG2, SK-HEP1, A-549, NCI-H146, NCI-H82, NCI-H82, SK-LU-1, WI-38, MRC-5, HLF-a, CCD-19Lu, C39, Hs294T, SK-MEL5, COLO 829, U266B1, RPMI 2650, BeWo, JEG-3, JAR, SW 1353, MeKam, and SCC-4; and higher plant cells. Appropriate culture media and conditions for the above-described host cells are known in the art.

[0316] The supernatants are collected and the protein production tested by Western blot or ELISA. The relative expression of the wild type and codon optimized constructs are compared.

[0317] In addition to plasmids encoding single SARS-CoV proteins, single plasmids which contain a portion of a SARS-CoV coding region are constructed according to standard methods. For example, portions of a SARS-CoV coding region that is too large to be contained in a single plasmid may be inserted into two or more plasmids. Also, single plasmids which contain two or more SARS-CoV coding regions are constructed according to standard methods. For example, a polycistronic construct, where two or more SARS-CoV coding regions are transcribed as a single transcript in eukaryotic cells may be constructed by sepa-

rating the various coding regions with IRES sequences (Jang et al. "A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation." *J. Virol.* 62: 2636-43 (1988); Jang et al. "Cap-independent Translation of Picornavirus RNAs: Structure and Function of the Internal Ribosomal Entry Site." *Enzyme* 44:292-309(1990).

[0318] Alternatively, two or more coding regions may be inserted into a single plasmid, each with their own promoter sequence.

Example 2

In Vitro Expression of SARS-CoV Subunit Proteins

[0319] Expression of SARS-CoV Nucleocapsid (N) and Spike (S) constructs were tested in vitro by transfection of a mouse melanoma cell line (VM92). The following expression constructs were transfected individually into VM92 cells and cultured for a period of time. All SARS-CoV sequences described below, were cloned into the VR1012 expression vector. The VR9208 expression plasmid contains a nucleotide sequence encoding the SARS-CoV S1 domain which was codon-optimized according to the full optimization method described herein and is disclosed in SEQ ID NO:50. The VR9204 expression plasmid contains a nucleotide sequence encoding a fragment of the SARS-CoV S1 which corresponds to amino acids 1-417 of the SARS-CoV S1 protein. The coding sequence in VR9204 was also codon optimized according to the full optimization method described herein.

[0320] VR9219—expressing full-length SARS-CoV N protein

[0321] VR9208—expressing SARS-CoV S1 domain of the S protein (amino acids 1-683 of the S protein)

[0322] VR9204—expressing a fragment of the SARS-CoV S1 domain (amino acids 1-417 of the S1 domain)

[0323] VR9209—expressing SARS-CoV S2 domain of the S protein

[0324] VR9210—expressing SARS-CoV secreted S protein

[0325] Both cell extracts and cell culture medium supernatants were analyzed by Western blot. The presence of the SARS-CoV N protein and S proteins were detected using commercial rabbit polyclonal antibodies which recognize the N protein from SARS-CoV strain Urbani (IMG-543; Imgenex, San Diego, Calif) and the S proteins from SARS-CoV strain Urbani (IMG-557, 542 and 541; Imgenex, Diego, Calif). Western blot results are summarized below:

[0326] In both the supernatant and cell lysates from cells transfected with the VR9219 plasmid, protein bands of a molecular weight of between 37 and 50 kDa (as estimated by a protein molecular weight standard) were detectable. The SARS-CoV N protein has an expected molecule weight of 46 kDa. This result is consistent with efficient expression of the SARS-CoV N antigen.

[0327] The supernatant and cell lysates from cells transfected with four different SARS-CoV S antigen constructs were individually analyzed for the presence of the S antigen. The results are summarized below.

[0328] A protein band of 85-110 kDa (as estimated by a protein molecular weight standard) was detected by Western blot in both the lysate and supernatant of cells transfected with the VR9204 plasmid (S1 domain—fragment).

[0329] A protein band of about 150 kDa (as estimated by a protein molecular weight standard) was detected by Western blot in both the lysate and supernatant of cells transfected with the VR9208 plasmid (S1 domain).

[0330] A protein band of approximately 111 kDa (as estimated by a protein molecular weight standard) was detected by Western blot in both the lysate and supernatant of cells transfected with the VR9209 plasmid (S2 domain).

[0331] A protein band of about 190 kDa (as estimated by a protein molecular weight standard) was detected by Western blot in both the lysate and supernatant of cells transfected with the VR9210 plasmid (secreted S).

[0332] These results are consistent with efficient expression and secretion of SARS-CoV Spike protein. Due to the presence of glycosylation sites in the S protein, the molecular weight is difficult to accurately predict.

Example 3

Preparation of SARS-CoV Subunit Proteins

[0333] Recombinantly prepared SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, for use as subunit proteins in the various combination therapies and compositions described herein, are prepared using the following procedure.

[0334] Eukaryotic cells transfected with expression plasmids such as those described in Example 1 are used to express SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg. Alternatively, a baculovirus system can be used wherein insect cells such as, but not limited to, Sf9, Sf21, or D.Mel-2 cells are infected with recombinant baculoviruses which can express SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg. Other *in vitro* expression systems may be used, and are well known to those of ordinary skill in the art. For baculovirus expression of non-secreted forms of these proteins, cells which are infected with recombinant baculoviruses capable of expressing SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, are collected by knocking and scraping cells off the bottom of the flask in which they are grown. Cells infected with baculoviruses for 24 or 48 hours are less easy to detach

from flask and may lyse, thus care must be taken with their removal. Eukaryotic cells which are transfected, either transiently or permanently, with expression plasmids encoding non-secreted forms of SARS-CoV proteins are gently scraped off the bottom of the flasks in which they are grown. Flasks containing the cells are then rinsed with PBS and the cells are transferred to 250 ml conical tubes. The tubes are spun at 1000 rpm in J-6 centrifuge (300xg) for about 5-10 minutes. The cell pellets are washed two times with PBS and then resuspended in about 10-20 ml of PBS in order to count. The cells are finally resuspended at a concentration of about 2×10^7 cells/ml in RSB (10 mM Tris pH=7.5, 1.5 mM $MgCl_2$, 10 mM KCl).

[0335] At this point either a total cell lysate is prepared, or cytoplasmic and nuclear fractions are separated. Approximately 10^6 infected cells are used per lane of a standard SDS-PAGE mini-protein gel for gel analysis purposes. When separating cytoplasmic and nuclear fractions, 10% NP40 is added to the cells for a final concentration of 0.5%. The cell-NP40 mixture is vortexed and placed on ice for 10 minutes, vortexing occasionally. After ice incubation, the cells are spun at 1500 rpm in a J-6 centrifuge (600x) for 10 minutes. The supernatant is removed, which is the cytoplasmic fraction. The remaining pellet, containing the nuclei, is washed two times with buffer C (20 mM HEPES pH=7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSE, 0.5 mM DTT) to remove cytoplasmic proteins. The nuclei are resuspended in buffer C to 5×10^7 nuclei/ml. The nuclei are vortexed vigorously to break up particles and an aliquot is removed for the mini-protein gel, which is the nuclei fraction.

[0336] Whole cell lysates are prepared by simply resuspending the requisite number of cells in gel sample buffer.

[0337] For gel analysis, a small amount (about 10^6 nuclear equivalents) of the nuclear pellet is resuspended directly in gel sample buffer and run with equivalent amounts of whole cells, cytoplasm, and nuclei. Those fractions containing the SARS-CoV protein of interest are detected by Western blot analysis as described herein.

[0338] Following analysis as described above, larger quantities of crude subunit proteins are prepared from batch cell cultures by protein purification methods well known to those of ordinary skill in the art, e.g., the use of HPLC.

[0339] Secreted versions of SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, are isolated from cell culture supernatants using various protein purification methods well known to those of ordinary skill in the art.

Example 4

Preparation of Vaccine Formulations

[0340] Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either

alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are formulated with the poloxamer CRL 1005 and BAK (Benzalkonium chloride 50% solution, available from Ruger Chemical Co. Inc.) by the following methods. Specific final concentrations of each component of the formulae are described in the following methods, but for any of these methods, the concentrations of each component may be varied by basic stoichiometric calculations known by those of ordinary skill in the art to make a final solution having the desired concentrations.

[0341] For example, the concentration of CRL 1005 is adjusted depending on, for example, transfection efficiency, expression efficiency, or immunogenicity, to achieve a final concentration of between about 1 mg/ml to about 75 mg/ml, for example, about 1 mg/ml, about 2 mg/ml, about 3 mg/ml, about 4 mg/ml, about 5 mg/ml, about 6.5 mg/ml, about 7 mg/ml, about 7.5 mg/ml, about 8 mg/ml, about 9 mg/ml, about 10 mg/ml, about 15 mg/ml, about 20 mg/ml, about 25 mg/ml, about 30 mg/ml, about 35 mg/ml, about 40 mg/ml, about 45 mg/ml, about 50 mg/ml, about 55 mg/ml, about 60 mg/ml, about 65 mg/ml, about 70 mg/ml, or about 75 mg/ml of CRL 1005.

[0342] Similarly, the concentration of DNA is adjusted depending on many factors, including the amount of a formulation to be delivered, the age and weight of the subject, the delivery method and route and the immunogenicity of the antigen being delivered. In general, formulations of the present invention are adjusted to have a final concentration from about 1 ng/ml to about 30 mg/ml of plasmid (or other polynucleotide). For example, a formulation of the present invention may have a final concentration of about 1 ng/ml, about 5 ng/ml, about 10 ng/ml, about 50 ng/ml, about 100 ng/ml, about 500 ng/ml, about 1 µg/ml, about 5 µg/ml, about 10 µg/ml, about 50 µg/ml, about 200 µg/ml, about 400 µg/ml, about 600 µg/ml, about 800 µg/ml, about 1 mg/ml, about 2 mg/ml, about 2.5, about 3 mg/ml, about 3.5, about 4 mg/ml, about 4.5, about 5 mg/ml, about 5.5 mg/ml, about 6 mg/ml, about 7 mg/ml, about 8 mg/ml, about 9 mg/ml, about 10 mg/ml, about 20 mg/ml, or about 30 mg/ml of a plasmid.

[0343] Certain formulations of the present invention include a cocktail of plasmids (see, e.g., Example 1 *supra*) of the present invention, e.g., comprising coding regions encoding SARS-CoV proteins, for example SARS-CoV S, S1, S2, N, M, or E and optionally, plasmids encoding immunity enhancing proteins, e.g., cytokines. Various plasmids desired in a cocktail are combined together in PBS or other diluent prior to the addition to the other ingredients. Furthermore, plasmids may be present in a cocktail at equal proportions, or the ratios may be adjusted based on, for example, relative expression levels of the antigens or the relative immunogenicity of the encoded antigens. Thus, various plasmids in the cocktail may be present in equal proportions, or up to twice or three times as much of one plasmid may be included relative to other plasmids in the cocktail.

[0344] Additionally, the concentration of BAK may be adjusted depending on, for example, a desired particle size and improved stability. Indeed, in certain embodiments, formulations of the present invention include CRL 1005 and DNA, but are free of BAK. In general BAK-containing

formulations of the present invention are adjusted to have a final concentration of BAK from about 0.05 mM to about 0.5 mM. For example, a formulation of the present invention may have a final BAK concentration of about 0.05 mM, 0.1 mM, 0.2 mM, 0.3 mM, 0.4 mM, or 0.5 mM.

[0345] The total volume of the formulations produced by the methods below may be scaled up or down, by choosing apparatus of proportional size. Finally, in carrying out any of the methods described below, the three components of the formulation, BAK, CRL 1005, and plasmid DNA, may be added in any order. In each of these methods described below the term "cloud point" refers to the point in a temperature shift, or other titration, at which a clear solution becomes cloudy, i.e., when a component dissolved in a solution begins to precipitate out of solution.

Thermal Cycling of a Pre-Mixed Formulation

[0346] This example describes the preparation of a formulation comprising 0.3 mM BAK, 7.5 mg/ml CRL 1005, and 5 mg/ml of DNA in a total volume of 3.6 ml. The ingredients are combined together at a temperature below the cloud point and then the formulation is thermally cycled to room temperature (above the cloud point) several times, according to the protocol outlined in FIG. 2.

[0347] A 1.28 mM solution of BAK is prepared in PBS, 846 µl of the solution is placed into a 15 ml round bottom flask fitted with a magnetic stirring bar, and the solution is stirred with moderate speed, in an ice bath on top of a stirrer/hotplate (hotplate off) for 10 minutes. CRL 1005 (27 µl) is then added using a 100 µl positive displacement pipette and the solution is stirred for a further 60 minutes on ice. Plasmids comprising coding regions or codon-optimized coding regions encoding SARS-CoV proteins, for example, S, S1, S2, N, M, or E, as described herein, and optionally, additional plasmids comprising codon-optimized or non-codon-optimized coding regions encoding, e.g., additional SARS-CoV proteins, and/or other proteins, e.g., cytokines, are mixed together at desired proportions in PBS to achieve 6.4 mg/ml total DNA. This plasmid cocktail is added dropwise, slowly, to the stirring solution over 1 min using a 5 ml pipette. The solution at this point (on ice) is clear since it is below the cloud point of the poloxamer and is further stirred on ice for 15 min. The ice bath is then removed, and the solution is stirred at ambient temperature for 15 minutes to produce a cloudy solution as the poloxamer passes through the cloud point.

[0348] The flask is then placed back into the ice bath and stirred for a further 15 minutes to produce a clear solution as the mixture is cooled below the poloxamer cloud point. The ice bath is again removed and the solution stirred at ambient temperature for a further 15 minutes. Stirring for 15 minutes above and below the cloud point (total of 30 minutes), is defined as one thermal cycle. The mixture is cycled six more times. The resulting formulation may be used immediately, or may be placed in a glass vial, cooled below the cloud point, and frozen at -80° C. for use at a later time.

Thermal Cycling, Dilution and Filtration of a Pre-mixed Formulation, Using Increased Concentrations of CRL 1005

[0349] This example describes the preparation of a formulation comprising 0.3 mM BAK, 34 mg/ml or 50 mg/ml CRL 1005, and 2.5 mg/ml of DNA in a final volume of 4.0 ml. The ingredients are combined together at a temperature

below the cloud point, then the formulation is thermally cycled to room temperature (above the cloud point) several times, diluted, and filtered according to the protocol outlined in FIG. 3.

[0350] Plasmids comprising wild-type or codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, and/or other proteins, e.g., cytokines, are mixed together at desired proportions in PBS to achieve 6.4 mg/ml total DNA. This plasmid cocktail is placed into the 15 ml round bottom flask fitted with a magnetic stirring bar, and for the formulation containing 50 mg/ml CRL 1005, 3.13 ml of a solution containing about 3.2 mg/ml of e.g., S1 encoding plasmid and about 3.2 mg/ml S2 encoding plasmid (about 6.4 mg/ml total DNA) is placed into the 15 ml round bottom flask fitted with a magnetic stirring bar, and the solutions are stirred with moderate speed, in an ice bath on top of a stirrer/hotplate (hotplate off) for 10 minutes. CRL 1005 (136 μ l for 34 mg/ml final concentration, and 100 μ l for 50 mg/ml final concentration) is then added using a 200 μ l positive displacement pipette and the solution is stirred for a further 30 minutes on ice. Solutions of 1.6 mM and 1.8 mM BAK are prepared in PBS, and 739 μ l of 1.6 mM and 675 μ l of 1.8 mM are then added dropwise, slowly, to the stirring poloxamer solutions with concentrations of 34 mg/ml or 50 mg/ml mixtures, respectively, over 1 min using a 1 ml pipette. The solutions at this point are clear since they are below the cloud point of the poloxamer and are stirred on ice for 30 min. The ice baths are then removed; the solutions stirred at ambient temperature for 15 minutes to produce cloudy solutions as the poloxamer passes through the cloud point.

[0351] The flasks are then placed back into the ice baths and stirred for a further 15 minutes to produce clear solutions as the mixtures cooled below the poloxamer cloud point. The ice baths are again removed and the solutions stirred for a further 15 minutes. Stirring for 15 minutes above and below the cloud point (total of 30 minutes), is defined as one thermal cycle. The mixtures are cycled two more times.

[0352] In the meantime, two SteriLip® 50 ml disposable vacuum filtration devices, each with a 0.22 μ m Millipore Express® membrane (available from Millipore, cat # SCGP00525) are placed in an ice bucket, with a vacuum line attached and left for 1 hour to allow the devices to equilibrate to the temperature of the ice. The poloxamer formulations are then diluted to 2.5 mg/ml DNA with PBS and filtered under vacuum.

[0353] The resulting formulations may be used immediately, or may be transferred to glass vials, cooled below the cloud point, and frozen at -80° C. for use at a later time.

A Simplified Method Without Thermal Cycling

[0354] This example describes a simplified preparation of a formulation comprising 0.3 mM BAK, 7.5 mg/ml CRL 1005, and 5 mg/ml of DNA in a total volume of 2.0 ml. The ingredients are combined together at a temperature below the cloud point and then the formulation is simply filtered and then used or stored, according to the protocol outlined in FIG. 4.

[0355] A 0.77 mM solution of BAK is prepared in PBS, and 780 μ l of the solution is placed into a 15 ml round bottom flask fitted with a magnetic stirring bar, and the solution is stirred with moderate speed, in an ice bath on top of a stirrer/hotplate (hotplate off) for 15 minutes. CRL 1005 (15 μ l) is then added using a 100 μ l positive displacement pipette and the solution is stirred for a further 60 minutes on ice. Plasmids comprising coding regions or codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, and/or other proteins, e.g., cytokines, are mixed together at desired proportions in PBS to achieve a final concentration of about 6.3 mg/ml total DNA. This plasmid cocktail is added dropwise, slowly, to the stirring solution over 1 min using a 5 ml pipette. The solution at this point (on ice) is clear since it is below the cloud point of the poloxamer and is further stirred on ice for 15 min.

[0356] In the meantime, one SteriLip® 50 ml disposable vacuum filtration device, with a 0.22 μ m Millipore Express® membrane (available from Millipore, cat # SCGP00525) is placed in an ice bucket, with a vacuum line attached and left for 1 hour to allow the device to equilibrate to the temperature of the ice. The poloxamer formulation is then filtered under vacuum, below the cloud point and then allowed to warm above the cloud point. The resulting formulations may be used immediately, or may be transferred to glass vials, cooled below the cloud point and then frozen at -80° C. for use at a later time.

Example 5

Animal Immunizations

[0357] The immunogenicity of the various SARS-CoV expression products encoded polynucleotides and codon-optimized polynucleotides described herein are initially evaluated based on each plasmid's ability to mount an immune response in vivo. Plasmids are tested individually and in combinations by injecting single constructs as well as multiple constructs. Immunizations are initially carried out in animals, such as mice, rabbits, goats, sheep, domestic cats, non-human primates, or other suitable animal, by intramuscular (IM) injections. Serum is collected from immunized animals, and the antigen specific antibody response is quantified by ELISA assay using purified immobilized antigen proteins in a protein-immunized subject antibody—anti-species antibody type assay, according to standard protocols. The tests of immunogenicity further include measuring antibody titer, neutralizing antibody titer, T-cell proliferation, T-cell secretion of cytokines, and cytolytic T cell responses. Correlation to protective levels of the immune responses in humans are made according to methods well known by those of ordinary skill in the art. See above.

A. DNA Formulations

[0358] Plasmid DNA is formulated with a poloxamer by any of the methods described in Example 3. Alternatively, plasmid DNA is prepared as described above and dissolved at a concentration of about 0.1 mg/ml to about 10 mg/ml,

preferably about 1 mg/ml, in PBS with or without transfection-facilitating cationic lipids, e.g., DMRIE/DOPE at a 4:1 DNA:lipid mass ratio. Alternative DNA formulations include 150 mM sodium phosphate instead of PBS, adjuvants, e.g., Vaxfectin™ at a 4:1 DNA: Vaxfectin™ mass ratio, mono-phosphoryl lipid A (detoxified endotoxin) from *S. minnesota* (MPL) and trehalosuccinylmonoycolate AF (TDM), in 2% oil (squalene)-Tween 80-water (MPL+TDM, available from Sigma/Aldrich, St. Louis, Mo., catalog # M6536), a solubilized mono-phosphoryl lipid A formulation (AF, available from Corixa), or (α)-N-(3-Acetoxypropyl)-N,N-dimethyl-2,3-bis(octyloxy)-1-propanaminium chloride (compound # VC1240) (see Shriver, J. W. et al., *Nature* 415:331-335 (2002), and P.C.T. Publication No. WO 02/0084 A2, each of which is incorporated herein by reference in its entirety).

B. Animal Immunizations

[0359] Plasmid constructs comprising codon-optimized or non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are injected into BALB/c mice as single plasmids or as cocktails of two or more plasmids, as either DNA in PBS or formulated with the polyplex-based delivery system: 2 mg/ml DNA, 3 mg/ml CLX 1005, and 0.1 mM BAK. Groups of 10 mice are immunized three times, at biweekly intervals, and serum is obtained to determine antibody titers to each of the antigens. Groups are also included in which mice are immunized with a trivalent preparation, containing each of three plasmid constructs expressing any of the SARS-CoV polypeptides, e.g., soluble, extracellular S1, M, and N polypeptides, in equal mass.

[0360] An example of an immunization schedule is as follows:

Day -3	Pre-bleed
Day 0	Plasmid injections, intramuscular, bilateral in rectus femoris, 5-50 µg/leg
Day 20	Serum Collection
Day 21	Plasmid injections, intramuscular, bilateral in rectus femoris, 5-50 µg/leg
Day 48	Serum Collection
Day 49	Plasmid injections, intramuscular, bilateral in rectus femoris, 5-50 µg/leg
Day 59	Serum collection

[0361] Serum antibody titers, at the various time points are determined by ELISA, using as the antigen SARS-CoV protein preparations including, but not limited to, purified recombinant proteins, transfection supernatants and lysates from mammalian or insect cells transfected with the various plasmids described herein, or live, inactivated, or lysed SARS-CoV virus.

C. Immunization of Mice with Vaccine Formulations Using a VAXFECTIN™ Adjuvant

[0362] VAXFECTIN™ (a 1:1 molar ratio of the cationic lipid VC1052 and the neutral co-lipid DpPE) is a synthetic

cationic lipid formulation which has shown promise for its ability to enhance antibody titers against an antigen when administered with DNA encoding the antigen intramuscularly to mice. See Hartikka et al. "Vaxfectin Enhances the Humoral Response to Plasmid DNA-encoded Antigens." *Vaccine* 19: 1911-1923 (2001).

[0363] In mice, intramuscular injection of VAXFECTIN™ formulated with, for example, DNA encoding the IAV NP protein increased antibody titers to NP up to 20-fold to levels that could not be reached with DNA alone. In rabbits, complexing DNA with VAXFECTIN™ enhanced antibody titers up to 50-fold. Thus, VAXFECTIN™ shows promise as a delivery system and as an adjuvant in a DNA vaccine.

[0364] Vaxfectin mixtures are prepared by mixing chloroform solutions of VC1052 cationic lipid with chloroform solutions of DpPE neutral co-lipid. Dried films are prepared in 2 ml sterile glass vials by evaporating the chloroform under a stream of nitrogen, and placing the vials under vacuum overnight to remove solvent traces. Each vial contains 1.5 µmole each of VC1052 and DpPE. Liposomes are prepared by adding sterile water followed by vortexing. The resulting liposome solution is mixed with DNA at a phosphate mole:cationic lipid mole ratio of 4:1.

[0365] Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are mixed together at desired proportions in PBS to achieve a final concentration of at 1.0 mg/ml. The plasmid cocktail, as well as the controls, are formulated with VAXFECTIN™. Groups of 5 Balb/c female mice are injected bilaterally in the rectus femoris muscle with 50 µl of DNA solution (100 µl total/mouse), on days 1 and 21 and 49 with each formulation. Mice are bled for serum on days 0 (prebleed), 20 (bleed 1), and 41 (bleed 2), and 62 (bleed 3), and up to 4 weeks post-injection. Antibody titers to the various SARS CoV proteins encoded by the plasmid DNAs are measured by ELISA as described elsewhere herein.

[0366] Cytolytic T-cell responses are measured as described in Hartikka et al. "Vaxfectin Enhances the Humoral Response to Plasmid DNA-encoded Antigens." *Vaccine* 19: 1911-1923 (2001) and is incorporated herein in its entirety by reference. Standard ELISPOT technology is used for the CD4+ and CD8+ T-cell assays as described in Example 6, part A.

D. Production of SARS-CoV Antisera in Animals

[0367] Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are prepared according to the immunization scheme described above and injected into a suitable animal for generating polyclonal antibodies. Serum is collected and the antibody titered as above.

[0368] Monoclonal antibodies are also produced using hybridoma technology. Kohler, et al., *Nature* 256:495 (1975); Kohler, et al., *Eur. J. Immunol.* 6:511 (1976); Kohler, et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling, et al., in *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., (1981), pp. 563-681, each of which is incorporated herein by reference in its entirety. In general, such procedures involve immunizing an animal (preferably a mouse) as described above. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (Sp20), available from the American Type Culture Collection, Rockville, Md. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al., *Gastroenterology* 80:225-232 (1981), incorporated herein by reference in its entirety. The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the various SARS-CoV proteins.

[0369] Alternatively, additional antibodies capable of binding to SARS-CoV proteins described herein may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, various SARS-CoV-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the SARS-CoV protein-specific antibody can be blocked by the cognate SARS-CoV protein. Such antibodies comprise anti-idiotypic antibodies to the SARS-CoV protein-specific antibody and can be used to immunize an animal to induce formation of further SARS-CoV-specific antibodies.

[0370] It will be appreciated that Fab and F(ab)₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as pepsin (to produce Fab fragments) or pepsin (to produce F(ab)₂ fragments). Alternatively, SARS-CoV polypeptide binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

[0371] It may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. See, for review, Morrison, *Science* 229:1202 (1985); Oi, et al., *BioTechniques* 4:214 (1986); Cabilly, et al., U.S. Pat. No. 4,816,567; Taniguchi, et al., EP 171496; Morrison, et al., EP 173494; Neuberger, et al., WO 8601533; Robinson, et al., WO 8702671; Boulianne, et al., *Nature* 312:643 (1984); Neuberger, et al., *Nature* 314:268 (1985).

[0372] These antibodies are used, for example, in diagnostic assays, as a research reagent, or to further immunize animals to generate SARS-CoV-specific anti-idiotypic antibodies. Non-limiting examples of uses for anti-SARS-CoV

antibodies include use in Western blots, ELISA (competitive, sandwich, and direct), immunofluorescence, immunoelectron microscopy, radioimmunoassay, immunoprecipitation, agglutination assays, immunodiffusion, immunoelectrophoresis, and epitope mapping. Weir, D. *Ed. Handbook of Experimental Immunology*, 4th ed. Vols. I and II, Blackwell Scientific Publications (1986).

Example 6

Mouse and Rabbit Immunogenicity Studies to SARS-CoV Antigens

[0373] Balb/c mice were injected intramuscularly bilaterally with 100 µg of SARS-CoV antigen expressing plasmid. VR9204, VR9208, VR9209, VR9210, VR9219 plasmids were formulated in PBS and DMRIE:DOPE at a 4:1 DNA:lipid mass ratio.

[0374] New Zealand white rabbits were injected intramuscularly bilaterally with 1 mg of SARS-CoV antigen expressing plasmid (VR9219 (N antigen) or VR9204 (S1 fragment antigen)), formulated with DMRIE:DOPE, on days 1, 28 and 56. Rabbit sera anti-antigen titers were determined by ELISA assay. The ELISA assay was performed according to standard protocols. ELISA plates used in the assay were coated with cell culture supernatants, from cells transfected with the a SARS-CoV antigen plasmid. Sera from rabbits which had been injected with the corresponding plasmid was then applied to the plates. Bound rabbit antibodies were detected using an alkaline phosphatase-modified donkey anti-rabbit IgG monoclonal antibody (Jackson Immuno Research; Cat No. 711-055-152). Bound antibodies were detected by standard colorimetric method after 2.5 hours of incubation with chromogenic substrates. Optical Density was determined at a wavelength of 405 nm. The results of the ELISA assay are summarized below.

[0375] Data shown in Table 20 demonstrate the presence of anti-nucleocapsid antibodies at day 21 in rabbits injected with plasmid VR9219 expressing full-length SARS-CoV nucleocapsid antigen. The antibody titers reach a plateau at day 42 (1:400 dilution).

[0376] In another experiment, rabbits were injected with plasmid VR9204, which expresses a fragment of the SARS-CoV Spike S1 domain. ELISA plates were coated with in vitro-produced full length-secreted Spike protein from cells transfected with plasmid VR9210. Antibodies IMG-542 and IMG-557, which recognize amino acids 288-303 and 1124-1140 of the SARS-CoV spike protein respectively (available from Imgenex, San Diego, Calif.), were used as positive controls in the ELISA assay. An ELISA plate coated with supernatant from VR1012-transfected VM92 cells was used as a negative control in the ELISA assay. The data shown in Table 20 demonstrate the presence of anti-Spike antibodies at days 42 and 50 after injection.

TABLE 20

	Anti-SARS CoV Antigen Titers (Rabbits)	
	Nucleocapsid Plasmid - VR9219 1/500 sera dilution	S1 fragment Plasmid - VR9204 1/500 sera dilution
Day 21	0.92	0.22
Day 42	3.9	0.74
Day 50	NA	0.51
Day 80	4	NA

TABLE 20-continued

Anti-SARS CoV Antigen Titers (Rabbits)		
	Nucleocapsid Plasmid - VR9219 Vero sera dilution	S1 fragment Plasmid - VR9204 Vero sera dilution
Pre-bleed	0.13	0.19
IMG-542	NA	0.44
IMG-557	NA	2.41
VR1012	0.15	0.21

Example 7

Mucosal Vaccination and Electrically Assisted
Plasmid Delivery

A. Mucosal DNA Vaccination

[0377] Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S1, soluble S2, soluble TPA-S1, soluble TPA-S2, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, (100 µg/50 µl total DNA) are delivered to BALB/c mice at 0, 2 and 4 weeks via i.m., intranasal (i.n.), intravenous (i.v.), intravaginal (i.vag.), intrarectal (i.r.) or oral routes. The DNA is delivered unformulated, formulated with the cationic lipids DMRIE/DOPE (DD) or GAP-DMRIE/DOPE (GD), or formulated with a poloxamer as described in Example 3. As endpoints, serum IgG titers against the various SARS-CoV antigens are measured by ELISA and splenic T-cell responses are measured by antigen-specific production of IFN-γ and IL-4 in ELISPOT assays. Standard chromium release assays are used to measure specific cytotoxic T lymphocyte (CTL) activity against the various SARS-CoV antigens. In addition, IgG and IgA responses against the various SARS-CoV antigens are analyzed by ELISA of vaginal washes.

B. Electrically-Assisted Plasmid Delivery

[0378] In vivo gene delivery may be enhanced through the application of brief electrical pulses to injected tissues, a procedure referred to herein as electrically-assisted plasmid delivery. See, e.g., Aihara, H. & Miyazaki, J. *Nat. Biotechnol.* 16:867-70 (1998); Mir, L. M. et al., *Proc. Natl. Acad. Sci. USA* 96:4262-67 (1999); Hartikka, J. et al., *Mol. Ther.* 4:407-15 (2001); and Mir, L. M. et al.; Rizzuto, G. et al., *Hum Gene Ther* 11:1891-900 (2000); Widera, G. et al., *J. of Immunol.* 164: 4635-4640 (2000). The use of electrical pulses for cell electroporation has been used to introduce foreign DNA into prokaryotic and eukaryotic cells in vitro. Cell permeabilization can also be achieved locally, in vivo, using electrodes and optimal electrical parameters that are compatible with cell survival.

[0379] The electroporation procedure can be performed with various electroporation devices. These devices include external plate type electrodes or invasive needle/rod electrodes and can possess two electrodes or multiple electrodes placed in an array. Distances between the plate or needle

electrodes can vary depending upon the number of electrodes, size of target area and treatment subject.

[0380] The TriGrid needle array, used in examples described herein, is a three electrode array comprising three elongate electrodes in the approximate shape of a geometric triangle. Needle arrays may include single, double, three, four, five, six or more needles arranged in various array formations. The electrodes are connected through conductive cables to a high voltage switching device that is connected to a power supply.

[0381] The electrode array is placed into the muscle tissue, around the site of nucleic acid injection, to a depth of approximately 3 mm to 3 cm. The depth of insertion varies depending upon the target tissue and the size of the patient receiving electroporation. After injection of foreign nucleic acid, such as plasmid DNA, and a period of time sufficient for distribution of the nucleic acid, square wave electrical pulses are applied to the tissue. The amplitude of each pulse ranges from about 100 volts to about 1500 volts, e.g., about 100 volts, about 200 volts, about 300 volts, about 400 volts, about 500 volts, about 600 volts, about 700 volts, about 800 volts, about 900 volts, about 1000 volts, about 1100 volts, about 1200 volts, about 1300 volts, about 1400 volts, or about 1500 volts or about 1-1.5 kV/cm, based on the spacing between electrodes. Each pulse has a duration of about 1 µs to about 1000 µs, e.g., about 1 µs, about 10 µs, about 50 µs, about 100 µs, about 200 µs, about 300 µs, about 400 µs, about 500 µs, about 600 µs, about 700 µs, about 800 µs, about 900 µs, or about 1000 µs, and a pulse frequency on the order of about 1-10 Hz. The polarity of the pulses may be reversed during the electroporation procedure by switching the connectors to the pulse generator. Pulses are repeated multiple times. The electroporation parameters (e.g., voltage amplitude, duration of pulse, number of pulses, depth of electrode insertion and frequency) will vary based on target tissue type, number of electrodes used and distance of electrode spacing, as would be understood by one of ordinary skill in the art.

[0382] Immediately after completion of the pulse regimen, subjects receiving electroporation can be optionally treated with membrane stabilizing agents to prolong cell membrane permeability as a result of the electroporation.

[0383] Examples of membrane stabilizing agents include, but are not limited to, steroids (e.g., dexamethasone, methylprednisone and progesterone), angiotensin II and vitamin E. A single dose of dexamethasone, approximately 0.1 mg per kilogram of body weight, should be sufficient to achieve a beneficial affect.

[0384] EAPD techniques such as electroporation can also be used for plasmids contained in liposome formulations. The liposome—plasmid suspension is administered to the animal or patient and the site of injection is treated with a safe but effective electrical field generated, for example, by a TriGrid needle array. The electroporation may aid in plasmid delivery to the cell by destabilizing the liposome bilayer so that membrane fusion between the liposome and the target cellular structure occurs. Electroporation may also aid in plasmid delivery to the cell by triggering the release of the plasmid, in high concentrations, from the liposome at the surface of the target cell so that the plasmid is driven across the cell membrane by a concentration gradient via the pores created in the cell membrane as a result of the electroporation.

[0385] Female BALB/c mice aged 8-10 weeks are anesthetized with inhalant isoflurane and maintained under anesthesia for the duration of the electroporation procedure. The legs are shaved prior to treatment. Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are administered to BALB/c mice (n=10) via unilateral injection in the quadriceps with 25 µg total of a plasmid DNA per mouse using an 0.3 cc insulin syringe and a 26 gauge, ½ length needle fitted with a plastic collar to regulate injection depth. Approximately one minute after injection, electrodes are applied. Modified caliper electrodes are used to apply the electrical pulse. See Hartikka J. et al. *Mol Ther* 188:407-415 (2001). The caliper electrode plates are coated with conductivity gel and applied to the sides of the injected muscle before closing to a gap of 3 mm for administration of pulses. EAPD is applied using a square pulse type at 1-10 Hz with a field strength of 100-500 V/cm, 1-10 pulses, of 10-100 ms each.

[0386] Mice are vaccinated±EAPD at 0, 2 and 4 weeks. As endpoints, serum IgG titers against the various SARS-CoV antigens are measured by ELISA and splenic T-cell responses are measured by antigen-specific production of IFN-γ and IL-4 in ELISPOT assays. Standard chromium release assays are used to measure specific cytotoxic T lymphocyte (CTL) activity against the various SARS-CoV antigens.

[0387] Rabbits (n=3) are given bilateral injections in the quadriceps muscle with plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector. The implantation area is shaved and the TriGrid electrode array is implanted into the target region of the muscle. 3.0 mg of plasmid DNA is administered per dose through the injection port of the electrode array. An injection collar is used to control the depth of injection. Electroporation begins approximately one minute after injection of the plasmid DNA is complete. Electroporation is administered with a TriGrid needle array, with electrodes evenly spaced 7 mm apart, using an Ichor TGP-2 pulse generator. The array is inserted into the target muscle to a depth of about 1 to 2 cm. 4-8 pulses are administered. Each pulse has a duration of about 50-100 µs, an amplitude of about 1-1.2 kV/cm and a pulse frequency of 1 Hz. The injection and electroporation may be repeated.

[0388] Sera are collected from vaccinated rabbits at various time points. As endpoints, serum IgG titers against the various SARS-CoV antigens are measured by ELISA and PBMC T-cell proliferative responses are measured by antigen-specific production of IFN-γ and IL-4 in ELISPOT assays or by quantification of intracellular cytokine staining. Standard chromium release assays are used to

measure specific cytotoxic T lymphocyte (CTL) activity against the various SARS-CoV antigens.

[0389] To test the effect of electroporation on therapeutic protein expression in non-human primates, male or female rhesus monkeys are given either 2 or 6 EAPD-assisted i.m. injections of plasmid constructs comprising codon-optimized and/or non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, (0.1 to 10 mg DNA total per animal). Target muscle groups include, but are not limited to, bilateral rectus femoris, cranial tibialis, biceps, gastrocnemius or deltoid muscles. The target area is shaved and a needle array, comprising between 4 and 10 electrodes, spaced between 0.5-1.5 cm apart, is implanted into the target muscle. Once injections are complete, a sequence of brief electrical pulses is applied to the electrodes implanted in the target muscle using an Ichor TGP-2 pulse generator. The pulses have an amplitude of approximately 120-200V. The pulse sequence is completed within one second. During this time, the target muscle may make brief contractions or twitches. The injection and electroporation may be repeated.

[0390] Sera are collected from vaccinated monkeys at various time points. As endpoints, serum IgG titers against the various SARS-CoV antigens are measured by ELISA and PBMC T-cell proliferative responses are measured by antigen-specific production of IFN-γ and IL-4 in ELISPOT assays or by quantification of intracellular cytokine staining. Standard chromium release assays are used to measure specific cytotoxic T lymphocyte (CTL) activity against the various SARS-CoV antigens.

Example 8

Combinatorial DNA Vaccine Using Heterologous Prime-Boost Vaccination

[0391] This Example describes vaccination with a combinatorial formulation including one or more polynucleotides comprising at least one codon-optimized or non-codon optimized coding regions encoding a SARS-CoV protein or fragment, variant, or derivative thereof prepared with an adjuvant and/or transfection facilitating agent; and also an isolated SARS-CoV protein or fragment, variant, or derivative thereof. Thus, antigen is provided in two forms. The exogenous isolated protein stimulates antigen specific antibody and CD4+ T-cell responses, while the polynucleotide-encoded protein, produced as a result of cellular uptake and expression of the coding region, stimulates a CD8+ T-cell response. Unlike conventional "prime-boost" vaccination strategies, this approach provides different forms of antigen in the same formulation. Because antigen expression from the DNA vaccine doesn't peak until 7-10 days after injection, the DNA vaccine provides a boost for the protein component. Furthermore, the formulation takes advantage of the immunostimulatory properties of the bacterial plasmid DNA.

A. Formulation Determinations for SARS-CoV proteins

[0392] This example mainly describes this procedure using an S2 subunit protein; however, the methods described

herein are applicable to any SARS-CoV subunit protein combined with any polynucleotide vaccine formulation. For example any polynucleotide comprising a codon-optimized or non-codon-optimized coding region encoding any SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg may be combined with any subunit SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg. Because only a small amount of protein is needed in this method, it is conceivable that the approach could be used to reduce the dose of other types of protein or antibody based vaccines, not described herein, when administered in combination with the polynucleotides and polypeptides of the present invention. The decreased dosing of other vaccines would allow for the increased availability of scarce or expensive vaccines. This feature would be particularly important for vaccines against pandemic SARS or biological warfare agents.

[0393] In this example, an injection dose of 10 µg SARS-CoV S protein, subunit 2 (S2) DNA per mouse, prepared essentially as described in Example 2 and in Ulmer, J. B., et al., *Science* 259:1745-49 (1993) and Ulmer, J. B. et al., *J. Virol.* 72:5648-53 (1998) is pre-determined in dose response studies to induce T cell and antibody responses in the linear range of the dose response and results in a response rate of greater than 95% of mice injected. Each formulation, either a plasmid comprising a codon-optimized or non-codon-optimized coding region encoding S2 alone ("S2 DNA"), or S2 DNA+S2 protein formulated with Ribi I or the cationic lipids, DMRIE:DOPE or Vaxfectin, is prepared in the recommended buffer for that vaccine modality. For injections with S2 DNA formulated with cationic lipid, the DNA is diluted in 2xPBS to 0.2 mg/ml+/-purified recombinant S2 protein (produced in baculovirus as described in Example 2) at 0.08 mg/ml. Each cationic lipid is reconstituted from a dried film by adding 1 ml of sterile water for injection (SWFI) to each vial and vortexing continuously for 2 min., then diluted with SWFI to a final concentration of 0.15 mM. Equal volumes of S2 DNA (+/-S2 protein) and cationic lipid are mixed to obtain a DNA to cationic lipid molar ratio of 4:1. For injections with DNA containing Ribi I adjuvant (Sigma), Ribi I is reconstituted with saline to twice the final concentration. Ribi I (2x) is mixed with an equal volume of S2 DNA at 0.2 mg/ml in saline+/-S2 protein at 0.08 mg/ml. For immunizations without cationic lipid or Ribi, S2 DNA is prepared in 150 mM sodium phosphate buffer, pH 7.2. For each experiment, groups of 9 BALB/c female mice at 7-9 weeks of age are injected with 50 µl of S2 DNA+/-S2 protein, cationic lipid or Ribi I. Injections are given bilaterally in each rectus femoris at day 0 and day 21. The mice are bled by OSP on day 20 and day 33 and serum titers of individual mice are measured.

[0394] S2 specific serum antibody titers are determined by indirect binding ELISA using 96 well ELISA plates coated overnight at 4° C. with purified recombinant S2 protein at 0.5 µg per well in BBS buffer pH 8.3. S2-coated wells are blocked with 1% bovine serum albumin in BBS for 1 h at room temperature. Two-fold serial dilutions of sera in block-

ing buffer are incubated for 2 h at room temperature and detected by incubating with alkaline phosphatase conjugated (AP) goat anti-mouse IgG-Fc (Jackson ImmunoResearch, West Grove, Pa.) at 1:5000 for 2 h at room temperature. Color is developed with 1 mg/ml para-nitrophenyl phosphate (Calbiochem, La Jolla, Calif.) in 50 mM sodium bicarbonate buffer, pH 9.8 and 1 mM MgCl₂ and the absorbance read at 405 nm. The titer is the reciprocal of the last dilution exhibiting an absorbance value 2 times that of pre-bled samples.

[0395] Standard ELISPOT technology, used to identify the number of interferon gamma (IFN-γ) secreting cells after stimulation with specific antigen (spot forming cells per million splenocytes, expressed as SFU/million), is used for the CD4+ and CD8+ T-cell assays. For the screening assays, 3 mice from each group are sacrificed on day 34, 35, and 36. At the time of collection, spleens from each group are pooled, and single cell suspensions made in cell culture media using a dounce homogenizer. Red blood cells are lysed, and cells washed and counted. For the CD4+ and CD8+ assays, cells are serially diluted 3-fold, starting at 10⁶ cells per well and transferred to 96 well ELISPOT plates pre-coated with anti-murine IFN-γ monoclonal antibody. Spleen cells are stimulated with the H-2K^b binding peptide, TYQRTALV (SEQ ID NO: 55) at 1 µg/ml and recombinant murine IL-2 at 1 U/ml for the CD8+ assay and with purified recombinant S2 protein at 20 µg/ml for the CD4+ assay. Cells are stimulated for 20-24 hours at 37° C. in 5% CO₂, then the cells are washed out and biotin labeled anti-IFN-γ monoclonal antibody added for a 2 hour incubation at room temperature. Plates are washed and horseradish peroxidase-labeled avidin is added. After a 1-hour incubation at room temperature, AEC substrate is added and "spots" developed for 15 min. Spots are counted using the Immunospot automated spot counter (C.T.L. Inc., Cleveland Ohio.). Thus, CD4+ and CD8+ responses are measured in three separate assays, using spleens collected on each of three consecutive days.

B. Determining Combinatorial Formulations with SARS-CoV Polynucleotide Constructs

[0396] Plasmid constructs comprising codon-optimized or non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are used in the prime-boost compositions described herein. For the prime-boost modalities, the same protein may be used for the boost, e.g., DNA encoding S2 with S2 protein, or a heterologous boost may be used, e.g., DNA encoding S2 with an M protein boost. Each formulation, the plasmid comprising a coding region for the SARS-CoV protein alone, or the plasmid comprising a coding region for the SARS-CoV protein plus the isolated protein, is formulated with Ribi I or the cationic lipids, DMRIE:DOPE or Vaxfectin. The formulations are prepared in the recommended buffer for that vaccine modality. Exemplary formulations, using S2 as an example, are described herein. Other plasmid/protein formulations, including multivalent formulations, can be easily prepared by use of ordinary skill in the art by following this example. For injections with DNA formulated with cationic

lipid, the DNA is diluted in 2xPBS to 0.2 mg/ml +/- purified recombinant SARS-CoV protein at 0.08 mg/ml. Each cationic lipid is reconstituted from a dried film by adding 1 ml of sterile water for injection (SWFI) to each vial and vortexing continuously for 2 min., then diluted with SWFI to a final concentration of 0.15 mM. Equal volumes of S2 DNA (+/-S2 protein) and cationic lipid are mixed to obtain a DNA to cationic lipid molar ratio of 4:1. For injections with DNA containing Rib1 1 adjuvant (Sigma), Rib1 1 is reconstituted with saline to twice the final concentration. Rib1 1 (2x) is mixed with an equal volume of S2 DNA at 0.2 mg/ml in saline +/- S2 protein at 0.08 mg/ml. For immunizations without cationic lipid or Rib1, S2 DNA is prepared in 150 mM sodium phosphate buffer, pH 7.2. For each experiment, groups of 9 BALB/c female mice at 7-9 weeks of age are injected with 50 µl of S2 DNA +/- S2 protein, cationic lipid or Rib1 1. The formulations are administered to BALB/c mice (n=10) via bilateral injection in each rectus femoris at day 0 and day 21.

[0397] The mice are bled on day 20 and day 33, and serum titers of individual mice to the various SARS-CoV antigens are measured. Serum antibody titers specific for the various SARS-CoV antigens are determined by ELISA. Standard ELISPOT technology, used to identify the number of interferon gamma (IFN-γ) secreting cells after stimulation with specific antigen (spot forming cells per million splenocytes, expressed as SFU/million), is used for the CD4+ and CD8+ T-cell assays using 3 mice from each group vaccinated as above, sacrificed on day 34, 35, and 36, post vaccination.

Example 9

Challenge in Non-Human Primates

[0398] The purpose of these studies is to evaluate three or more of the optimal plasmid DNA vaccine formulations for immunogenicity in non-human primates. Preliminary challenge experiments may be carried out in other suitable animal modes, for example birds as described below, or in domestic cats, Rhesus or cynomolgus monkeys (6/group) are vaccinated with plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding SARS-CoV proteins, for example, SARS-CoV S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2 proteins, fusions thereof, or fragments, variants or derivatives of such proteins either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, intramuscularly 0.1 to 2 mg DNA combined with cationic lipid, and/or poloxamer and/or aluminum phosphate based or other adjuvants at 0, 1 and 4 months.

[0399] Blood is drawn twice at baseline and then again at the time of and two weeks following each vaccination, and then again 4 months following the last vaccination. At 2 weeks post-vaccination, plasma is analyzed for humoral response and PBMCs are monitored for cellular responses, by standard methods described herein. Animals are monitored for 4 months following the final vaccination to determine the durability of the immune response.

[0400] Animals are challenged within 2-4 weeks following the final vaccination. Animals are challenged intranasally with the suitable dose of virus based on preliminary challenge studies. Nasal swabs, pharyngeal swabs and lung

lavages are collected at days 0, 2, 4, 6, 8 and 11 post-challenge and will be assayed for cell-free virus titers on monkey kidney cells. After challenge, animals are monitored for clinical symptoms, e.g., rectal temperature, body weight, leukocyte counts, and in addition, hematocrit and respiratory rate. Oropharyngeal swab samples are taken to allow determination of the length of viral shedding. Illness is scored using a variety of conventional illness scoring methods such as the system developed by Berendts & Hall (*Infect Immun* 16:476-479 (1977)), and will be analyzed by analysis of variance and the method of least significant difference.

Example 10

Challenge in Birds

[0401] In this example, various vaccine formulations of the present invention are tested in a chicken SARS-CoV model. For these studies a SARS-CoV is used for the challenge. Plasmid constructs comprising codon-optimized and non-codon-optimized coding regions encoding S, S1, S2, N, M, E, soluble S, soluble S1, soluble S2, soluble TPA-S, soluble TPA-S1, and soluble TPA-S2, as described herein, fusions; or alternatively, coding regions (either codon-optimized or non-codon optimized) encoding various SARS-CoV proteins or fragments, variants or derivatives, either alone or as fusions with a carrier protein, e.g., HBcAg, as well as various controls, e.g., empty vector, are formulated with cationic lipid, and/or poloxamer and/or aluminum phosphate based or other adjuvants. The vaccine formulations are delivered at a dose of about 1-10 µg, delivered IM into the defeathered breast area, at 0 and 1 month. The animals are bled for antibody results 3 weeks following the second vaccine. Antibody titers against the various SARS-CoV antigens are determined using techniques described in the literature. See, e.g., Kodihalli S. et al., *Vaccine* 18:2592-9 (2000). The birds are challenged intranasally with 0.1 mL containing 100 LD₅₀ 3 weeks post second vaccination. The birds are monitored daily for 10 days for disease symptoms, which include gasping, coughing and nasal discharge, wet eyes and swollen sinuses, reduced food consumption and weight loss. Tracheal and cloacal swabs are taken 4 days following challenge for virus titration.

[0402] The present invention is not to be limited in scope by the specific embodiments described which are intended as single illustrations of individual aspects of the invention, and any compositions or methods which are functionally equivalent are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

[0403] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 69

<210> SEQ ID NO 1

<211> LENGTH: 3588

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 1

atgtttattt tcttattatt tcttactctc actagtggta gtgacctga coggtygacc 60
 acctttgatg atgttcaagg tcttaattac actcaacata ctctacttat geggygggtt 120
 tactatcctg atgaattttt tagatcagac actctttatt taactcagga tttatttctt 180
 ccattttatt ctaagttaac aggytttcat actattaato atactgttg caacctgtc 240
 atacacttta aggatgggtat ttattttgct gccacagaga aatcaaatgt tgtcctgggt 300
 tgggtttttg gttctacctt gaacaacaag tcacagtctg tgattattat taacaattct 360
 actaatgttg ttatacagag atgttaattt gaattgttg acaaccttt cttgtcgtt 420
 tctaaaccca tgggtacaca gacacatact atgatattcg ataagcatt taattgcact 480
 tctagatata tactctgatg cttttgctgt gatgttccag aagatcagg taattttaa 540
 caactacag agttttgtgt taaaaataaa gatgggttct totatgttta taaggcctat 600
 caactatag atgtagtctg tgactctact ctgtgtttta acactttgaa acatttttt 660
 aagtgtcttc ttgtatttaa cattacaatt tttagagcca ttcttcacgc cttttcaact 720
 gctcaagaca ttgggggcac gtcagctgca gctattttt tgggtattt aaagccaaact 780
 acatttatgo tcaagtatga tgaattgggt acaatcacag atgctgttga ttgttctcaa 840
 aatcaacttg ctgaactcaa atgctctgtt aagagctttg agattgacaa aggaatttac 900
 cagactctta atttcagggt tgttccctca ggagatgttg tgagattccc taattataca 960
 aaactgtgtc ctttttgaga ggtttttaat gtaactaat tccctctgtg ctatgatbg 1020
 gagagaaaaa aaattcttaa ttgtttgctt gattactctg tgcctacaaa ctcaacattt 1080
 ttctcaacct ttaagtgtta tggcgtttct gccactaagt tgaatgatct ttgtctctc 1140
 aatgtctatg cagattcttt ttagtctaat ggagatgat taagcaaaat agcgcaggga 1200
 caaactgggt ttattgtgta ttataattat aaattgccag atgatttcat gggttgtgtc 1260
 cttgtttgga atactagaaa cattgatgct acttcaactg gtaattataa ttataaatat 1320
 aggtatctta gacatggcaa gottaggccc tttagagag agtatcttaa ttggocttct 1380
 tccctctgatg gcaaaccttg caccacacct gctttaatt gttattggcc attaaatgat 1440
 tatgggtttt acaacctact tggcaattgg taacaacctt acagagtgt agtaccttct 1500
 ttigaacttt taagtgcac ggcacacggt ttgtggacaa aattatccac tgacctattt 1560
 aagaaccagt gtgtcaattt taattttaat ggactcaact gtaactgtgt gtttaactct 1620
 tcttcaaga gatctcaacc atttcaacaa ttggcctgt atgtttctga tttaactgat 1680
 tccgttctag atctcaaaac atctgaanta ttgaactttt cacttgctct ttttgggggt 1740
 gtaagtgtaa ttacacctgg acaaaatgct tcatctgaag ttgctgttct atataagat 1800
 gtaactgtca ctgatgttct tcaagcaatt catgcagatc aactcaacac agcttgggc 1860
 atatattcta ctggaacaaa tgaattccag actcaagcag gctgtcttat aggaactgag 1920

-continued

catgtgacaa cttcttatga gtgagacatt cctattggag ctggcatttg tgotagttaa 1980
 catacagitt ctttatatag tagtactaga caaaaatcta tbtggotta tactatgtct 2040
 ttaggtgtgt atagtcaat tgttactct aatacaacca tbtgtatacc tactaacatt 2100
 tcaattagca ttactacaga agtaatgoot gttcttatgt ctaaaaactc cgtagattgt 2160
 aatatgtaca tctggaggga ttctactgaa tbtgtaatt tgtctctcca atatggtgac 2220
 ttttgcacac aactaatctg tgcactctca ggtattgtct ctgaacagga tgcgaacaca 2280
 cgtgaagtgt tgcotcaagt caaacaagt tacaaaaacc caactttgaa atattttggt 2340
 ggttttaatt tttaacaaat attactgtac cctctaaaga caactaagag gtcttttatt 2400
 gaggacttgc tctttaataa ggtgacactc gctgatgtct gottactgaa gcaatatgac 2460
 gaatgcttag tbtatattaa tbttagagat ctcattttgt cgcagaagtt caatggactt 2520
 acagtgttgc caactctgtc cactgatgat atgatgtct cctaacctgc tgccttagtt 2580
 agtggtagct ccaactgtgt atggacatt ggtgtgtgt cgtctctcca aatacctttt 2640
 gctatgcata tggcatatag gttaactgac attggagta cccaatgtt tctctatgag 2700
 aacaaaaaac aaatgcocaa ccaatttaac aaggagatta gtcaaatca agaatcaact 2760
 acaacaacat caactgcatt ggcgaagtg caagaagtgt ttaacagaa tbtctaaaga 2820
 ttaaacacac tbttaaaaca acttagctct aattttgtgt caatttcaag tbtgtaaat 2880
 gatctccttt cgcagactga taaagttag ggcggagta aattgacag gtaataaca 2940
 ggcagacttc aaagccttca aaactatga acacaacac taactagggc tbtgtaact 3000
 agggcttctg ctactcttgc tbtactataa atgtctgagt gtgtcttttg acaatcaaaa 3060
 agagttgact ttttggaaa ggcctaacac cttatgtctt tccacaagc agcccgcat 3120
 ggtgttgtct tctactatgt cactatgtg ccatccagag agaggaactt caccacagcg 3180
 ccagcaattt gtcatgaag caaagcatac ttccctctgt aaggtgtttt tbtgttaatt 3240
 ggcactcttt gttttattac acagaggaac ttctttctc cacaataat tactacagac 3300
 aatacatttg tctcagaaa tbtgtatgtc gttattgcca tcaataaca caagtttat 3360
 gatctctctg aaactgagct cgaactatc aaagaagag tggacaagta cttcaaaaat 3420
 catacatcac cagatgtga tbtggcgac attcaaggca ttaacgcttc tbtctcaac 3480
 attcaaaaag aaattgacgc cctcaatgag gtgcataaaa atttaaatga atocactatt 3540
 gaactccaag aattgggaaa atatgaqcaa tatattaat ggccttgg 3588

<210> SEQ ID NO 2

<211> LENGTH: 1196

<212> TYPE: PRN

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 2

Met Phe Ile Phe Leu Leu Phe Thr Leu Thr Ser Gly Ser Asp Leu
 1 5 10 15
 Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
 20 25 30
 His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
 35 40 45
 Ser Asp Thr Leu Thr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
 50 55 60

-continued-

Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
 65 70 75 80
 Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
 85 90 95
 Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
 100 105 110
 Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
 115 120 125
 Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
 130 135 140
 Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
 145 150 155 160
 Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
 165 170 175
 Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly
 180 185 190
 Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
 195 200 205
 Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu
 210 215 220
 Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro
 225 230 235 240
 Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Ala Tyr Phe Val Gly Tyr
 245 250 255
 Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile
 260 265 270
 Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys
 275 280 285
 Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn
 290 295 300
 Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr
 305 310 315 320
 Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser
 325 330 335
 Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr
 340 345 350
 Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly
 355 360 365
 Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala
 370 375 380
 Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly
 385 390 395 400
 Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Phe
 405 410 415
 Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser
 420 425 430
 Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu
 435 440 445
 Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly
 450 455 460
 Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp

-continued

465	470	475	480
Tyr Gly Phe Tyr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val			
485		490	495
Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly			
500	505		510
Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn			
515	520		525
Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg			
530	535	540	
Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp			
545	550	555	560
Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys			
565	570		575
Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser			
580	585		590
Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr			
595	600		605
Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr			
610	615	620	
Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu			
625	630	635	640
His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile			
645	650		655
Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys			
660	665		670
Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala			
675	680		685
Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile			
690	695	700	
Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys			
705	710	715	720
Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu			
725	730		735
Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile			
740	745		750
Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys			
755	760		765
Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe			
770	775	780	
Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile			
785	790	795	800
Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met			
805	810		815
Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile			
820	825		830
Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr			
835	840	845	
Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala			
850	855	860	
Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe			
865	870	875	880

-continued

Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn
 885 890 895
 Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala
 900 905 910
 Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly
 915 920 925
 Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu
 930 935 940
 Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn
 945 950 955 960
 Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp
 965 970 975
 Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln
 980 985 990
 Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala
 995 1000 1005
 Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp
 1010 1015 1020
 Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala
 1025 1030 1035
 Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln
 1040 1045 1050
 Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys
 1055 1060 1065
 Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser
 1070 1075 1080
 Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr
 1085 1090 1095
 Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly
 1100 1105 1110
 Ile Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp
 1115 1120 1125
 Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser
 1130 1135 1140
 Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val
 1145 1150 1155
 Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys
 1160 1165 1170
 Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
 1175 1180 1185
 Glu Gln Tyr Ile Lys Trp Pro Trp
 1190 1195

<210> SEQ ID NO 3
 <211> LENGTH: 2049
 <212> TYPE: DNA
 <213> ORGANISM: SARS-CoV Urbani strain
 <400> SEQUENCE: 3

atgtttattt tcttattatt tcttactctc actagtggtta gtgaccttga cgggtgcacc 60
 acttttgatg atgttcaagc tctcaattac actcaacata cttactctat gagggggggt 120

-continued-

tactatctgt atgaatttt tagatcagac actctttatt taactcagga tttatttctt 180
 coattttatt ctaatgttao aggytttcat actattaato atactgtttg caacctgtc 240
 ataccttta aggatgggat ttattttgct gccacagaga aatcaaatgt tgcctgtggt 300
 tgggtttttg gttctacctat gaacacaaag tcacagtctg tgattattat taacaattct 360
 actaatgttg ttatacagag atgtaacttt gaattgtgtg aaacctcttt ctttgtgttt 420
 tctaaaccca tgggtgacaa gacacctact atgtattctg ataatgcatt taattgcact 480
 ttctagttaca tatctgatgc cttttgcttt gatgtttcag aaaaagtcagg taattttaaa 540
 caactacagag agtttgtgtt taaaataaaa gatgggttct tctatgttta taagggttat 600
 caacctatag atgtagtctt tgatctacct tctgttttta acctttgaa acctattttt 660
 aagtgtgctc ttggtattaa cattacaaat tttagagcca ttctttacagc cttttcaact 720
 gctcaagaca ttgggggcac gtcagtgtca gctatttttg ttgctatttt aaagccaaact 780
 acctttatgc tcaagtatga tgaaaatggt acaatccag atgtgtgtga ttgttctcaa 840
 aatcaacttg ctgaactcaa atgctctgtt aagagctttg agattgacaa aggaatttac 900
 cagaactcta atttcagggt tgttccctca ggaagtgttg tgagattccc taattattca 960
 aactgtgttc cttttggaga ggtttttaat gctactaaat tccctctgt ctatgcatgg 1020
 gagagaasaa aaatttctaa ttgtgtgtgt gattactctg tgcctacaa ctcaacttt 1080
 ttctcaacct ttaagtgtca tggcgtttct gccactaagt tgaatgatct ttgctctctc 1140
 aatgtctatg cagattcttt tttagtcaag ggaagtgtg taagcaaat agcgcacaga 1200
 caaacigtty ttattgtctg ttataattat aattgtccag atgaattcat ggggtgtgtc 1260
 ctgtcttggg atactaggaa cattgatgct acttcaactg gtaattataa ttataaatat 1320
 aggtatctta gacatggcaa gcttagccct tttagagag acatatctaa tgcgctttc 1380
 tccctgatg gcaaaccttg caccaccact gctcttaatt gttattggcc attaatgat 1440
 tatgttttt acacactac tggcattggo taccacact acagagttgt agtactttct 1500
 ttgtaacttt taatgcaccc ggcaacggtt tgtggaccaa aattatccac tgaacttatt 1560
 aagaaccagt gtgtcaattt taattttaat ggactcactg gtactgtgtg gttactctct 1620
 tcttcaaaag gatctcaaac atttcaacaa ttggcctgt atgtttctga ttcaactgat 1680
 tccgttcagag atctcaaaac atctgaata ttgaacttt cacttctctc ttttgggggt 1740
 gtaagtgtta ttacacttgg acaaaatgct tcatctgaag ttgtgtttct atatacagat 1800
 gttactgtca ctgatgttct tacagaact catgcagato aactcaaccc agcgtggcgc 1860
 atatatctta ctggaacaa ttgattccag actcaagcag gctgtcttat aggaagtgtg 1920
 catgtgcaca cttcttatga gtgcagactt cctattggag ctggcatttg tctagttaac 1980
 catcaagttt cttttataag tagtactaga caaaaatcta ttgtggctta tactattctc 2040
 ttaggtgct

<210> SEQ ID NO 4

<211> LENGTH: 693

<212> TYPE: PRN

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 4

Met Phe Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
 1 5 10 15

-continued

Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln	20	25	30
His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg	35	40	45
Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser	50	55	60
Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val	65	70	75
Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn	85	90	95
Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln	100	105	110
Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys	115	120	125
Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met	130	135	140
Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr	145	150	155
Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser	165	170	175
Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly	180	185	190
Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp	195	200	205
Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu	210	215	220
Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Thr	225	230	235
Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Ala Tyr Phe Val Gly Tyr	245	250	255
Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile	260	265	270
Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys	275	280	285
Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn	290	295	300
Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr	305	310	315
Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser	325	330	335
Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr	340	345	350
Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly	355	360	365
Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala	370	375	380
Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly	385	390	395
Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe	405	410	415

-continued-

Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser
420 425 430

Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu
435 440 445

Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly
450 455 460

Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp
465 470 475 480

Tyr Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val
485 490 495

Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly
500 505 510

Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn
515 520 525

Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg
530 535 540

Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp
545 550 555 560

Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys
565 570 575

Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser
580 585 590

Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr
595 600 605

Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr
610 615 620

Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu
625 630 635 640

His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile
645 650 655

Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys
660 665 670

Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala
675 680

<210> SEQ ID NO 5

<211> LENGTH: 1539

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 5

gatagttcaa ttgcttactc taataacacc attgctatca ctactaaactt ttcacttagc 60

attactacag aagtaagtcc tgtttctatg gctaaaacct ccgttagattg taatatgtac 120

atctgcggag attctactga atgtgctaat ttgcttctcc aatatgtagt cttttgcaca 180

caactaatc gtgcactctc aggtattgct gctgaacagc atcgaacac acgtgaagt 240

ttcgtccag taacaacaat gtacaaaacc ccaactttga aatattttgg tggttttaat 300

ttttcacaaa tattactcga cctctaaag ccaactaaga ggtcttttat tggagacttg 360

ctctttaata aggtgacctc cgtgatgct ggttctatga agcaatatgg cgaatgcta 420

ggtgatatta atgctagaga tctcatttgt gcgcagaagt tcaatggact tacagtgttg 480

ccactctgc tcaactgatga tatgattgct gctacactg ctgctctagt tagtggtaac 540

-continued

gccactgctg gatggacatt tgggtctggc gctgctcttc aaataccttt tgcctatgcaa 600
 atggcatata ggttcaatgg cattggagtt acccaaaatt ttctctatga gaaccaaaaa 660
 caaatcgcca accaatttaa caaggcgatt agtcaaatc aagatcaact tacaacaaca 720
 tcaactgcat tgggcaagct gcaagacgtt gtaaccaga atgctcaagc attaaacaca 780
 ctgtttaaac aacttagctc taattttggt gcaatttcaa gtgtgtctaaa tgaatctctt 840
 tcgcgacttg ataaagtctg ggcggaggta caaatggaca ggtaattac aggcgactt 900
 caaagccttc aaacctatgt aacacacaaa ctaatcggg ctgctgaaat cagggtctct 960
 gctaactctg ctgtacttaa aatgtctgag tgtgtctctg gacaatcaaa aagcgttgac 1020
 ttttgtggaa agggtaccca ccttatgtcc ttcccacaa cagcccgcga tgggttgttc 1080
 ttctacatg tcaagtatgt gccatccag gagaggaaat tcaccacagc gccgcgaatt 1140
 tgtcatgaag gcaagcata ctctccctgt gaaggtgttt ttgtgttaa tggcactctt 1200
 tggtttatta cacagaggaa ctctttttct ccacaataa ttactacaga caatcattt 1260
 gtctcaggaa attgtgtgtg cgttatggc atcattaaca accagttta tgaatctctg 1320
 caactgagc tgcactcatt caaagaaagc ctggacacgt acttcaaaaa tcatacatca 1380
 ccagatgttg atcttgccga catttcaggc attaacgtt ctgtgtctaa cattcaaaaa 1440
 gaaattgacc gcctcaatga ggtgcgtaaa aatttaaatg aatcaatcat tgaacttcaa 1500
 gaattgggaa aatgatgcca atatattaaa tggccttg 1539

<210> SEQ ID NO 6

<211> LENGTH: 513

<212> TYPE: PRN

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 6

Asp Ser Ser Ile Ala Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn
 1 5 10 15
 Phe Ser Ile Ser Ile Thr Thr Glu Val Met Pro Val Ser Met Ala Lys
 20 25 30
 Thr Ser Val Asp Cys Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys
 35 40 45
 Ala Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg
 50 55 60
 Ala Leu Ser Gly Ile Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val
 65 70 75 80
 Phe Ala Gln Val Lys Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe
 85 90 95
 Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr
 100 105 110
 Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala
 115 120 125
 Asp Ala Gly Phe Met Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn
 130 135 140
 Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu
 145 150 155 160
 Pro Pro Leu Leu Thr Asp Asp Met Ile Ala Ala Tyr Thr Ala Leu
 165 170 175

-continued

Val Ser Gly Thr Ala Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala
180 185 190

Leu Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile
195 200 205

Gly Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn
210 215 220

Gln Phe Asn Lys Ala Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr
225 230 235 240

Ser Thr Ala Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln
245 250 255

Ala Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile
260 265 270

Ser Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala
275 280 285

Glu Val Gln Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln
290 295 300

Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser
305 310 315 320

Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser
325 330 335

Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
340 345 350

Gln Ala Ala Pro His Gln Gly Val Val Phe Leu His Val Thr Val Pro
355 360 365

Ser Gln Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly
370 375 380

Lys Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser
385 390 395 400

Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr Thr
405 410 415

Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Ile
420 425 430

Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys
435 440 445

Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp
450 455 460

Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys
465 470 475 480

Glu Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu
485 490 495

Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro
500 505 510

Trp

<210> SEQ ID NO 7

<211> LENGTH: 3633

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 7

atggatgcaa tgaagagagg gctctgctgt gtctgctgc tgttgaggc agtctcggt 60
 togcaccagc cttagaggatc gggaaagtgc cttgaccggt gcaccacttt tgatgatgtt 120

-continued

caagctccta attacactca acatacttca tctatgaggg gggttacta tcoctatgaa	180
atitttagat cagacactct ttatttacct caggatttat ttctccatt ttattctaatt	240
gtacacgggt tctactatct taactatcag ttggcaacc ctgtctatacc tttaaaggat	300
ggatattatt ttgctgocac agagaactca aatgtgttcc gtggtgggt tttgtgtct	360
acactgaaca accaagtccca gtccgtgatt attattacca attctactaa tgttgtata	420
cagactgta actttgaatt gtgtgacaaa cttttcttgy ctgtttctaa acccaagggt	480
acacagacac atactatgat attogataat gcaatttaatt gcaatttoga gtacatatct	540
gatgccttt cgccttgatgt ttacagaagac tcaagtaatt ttaaacactt acagagattt	600
gtgtttaaaa ataagatgg gtctctctat gtttataagg gctatcaacc tatgatgta	660
gttctgtatc tacctcttgg tttaaacact ttgaacacta ttttaagtt gctcttgggt	720
attaacatta caaattttg agocattctt acagcctttt cactgtctca agacatttgg	780
ggcagctcag ctgcagccta tttgttggc tatttaaacg caactacatt tatgtctaa	840
tatgatgaaa atgtgtcaat cacagatgct gttgattgtt ctcaaatcc actgtctgaa	900
ctcaaatgct ctgttaagag ctttgagatt gacaaaggaa ttacacagac ctctaatttc	960
aggggtgtc ctccagagga ttgttgtaga ttccctaata ttcaaacact gtgtctttt	1020
ggagaggttt ttaagtctac taactccct ctgtctatg catgggagag aaaaaaat	1080
tcantgtgt ttgtgtatta ctctgtctc tacaactcaa cattttttt acactttaag	1140
tgctatggg tttctgocac taagtgaat gatcttggct tctcaaatgt ctatgcagat	1200
ctttttag tagaaggaga tgaagtata caaatagcag caggacaaac tgggtttatt	1260
gctgattata attataaatt gccagatgat tctactgggt gtgtcttgc ttggaact	1320
aggaacattg atgtacttc aactggtaatt tataattata aatatagga tcttagacat	1380
ggcaagctta ggcctttga gagagacata totaatgtgc ctctctccc tgatggcaa	1440
ccttgcaacc caoctgtct taattgttat tggcattaa atgattatgg tttttacac	1500
actactggca ttggctacaa accatacaga gtgtagtag ttcttttga acttttaaat	1560
gcacccggca cggtttggg accaaaatta tccactgaac ttattaaaga ccaagtgtgc	1620
aattttaatt ttaagtgaact caactgtact ggtgtgttaa ctctctctc aaagagattt	1680
caaccatttc aacaaatttg cctgtatgtt cctgatttca ctgattcctg tctagatcct	1740
aaaaatctg aatatattga catttcaact tgcctcttgg ggggtgtaag tgaattaca	1800
cctggaacaa atgtctcatc tgaagtgtct gttctatct aagatgttaa ctgcactgat	1860
gtttctacag caattctcgc agatcaact accacagctt ggccatata tctactgga	1920
aacaaatgat tccagactca agcaggctgt cttataggag ctgacatgt cgaactctt	1980
tatgtagtgc acattctcat tggagctggc atttgtccta gttacacata agttctotta	2040
ttactgtagta ctagccaaaa atctatttgy gcttatacta tgtcttttag tgcgtatgt	2100
tcaattgtt actctactaa caccattgct atactacta actttcaat tagcaactat	2160
acagaagtaa tgcctgttc tatggctaan accctcctag attgtaatat gtaactctgc	2220
ggagattcta ctgaatgtgc taatttgcct ctccaatatg gtatgtttg cacacaaata	2280
aactgtgac tctcaggtat tgcgtgcaa caggatgca acacacgta agtctctgt	2340
caagtcaaaa aaatgtacaa aaccccaact ttgaatatatt ttggtggtt taattttca	2400

-continued

caaatattac ctgaccctct aaagccaact aaggggtctt ttattgagga cttgctcttt 2460
 aataagggtga cactcgctga tgcggcttc atgaagcaat atggcggaatg cctaggtgat 2520
 ataatgtga gagatctcat ttgtgcgcag aagttcaatg gacttaacgt gttgcacact 2580
 ctgctcactg atgatatgat tgcgtgctac actgctgctc tagttagtgg taactgcacact 2640
 gctggatgga catttggtgc tgcgctgct cttonaatc cttttgctat gcaaatgcca 2700
 tatagggtta atggcattgg agttaccctaa aatgtctct atgagaacca aaacaaatc 2760
 gcaacccaat ttaacaaggc gattagtcaa attcaagaat cacttacaac aacatcaact 2820
 gcattggcca agctgcacga cgttgtaac cagaatgctc aagcattaaa cacacttgtt 2880
 aaacaactta gctctaattt tgggtcaatt tcaagtgtgc taatgatata cctttcgaga 2940
 cttgataaag tcgaggcgga ggtacaatt gacaggttaa ttacaggcag acttcaaaagc 3000
 cttaaacact atgtaacaca acaactaatc agggctgctg aaatcagggc ttctgtaaat 3060
 cttgctgcta ctaaaatgtc tgaagtgtgt cttggacact caaaagaggt tgacttttgt 3120
 ggaagggtct accacattat gtctctocca caagcagccc cgcagtgtgt tgcctctcta 3180
 catgtcacgt atgtgcctac ccaggagagg aacttaacca cagcgccagc aatttgtcat 3240
 gaaggcaaa gatactctcc tcgtgaaggt gtttttgtgt ttaatggcag ttcttggttt 3300
 attcacacga ggaactctt ttctccacaa ataattacta cagacatcac atttgtctca 3360
 ggaatttgtg atgtcgttat tggcatcatt acaacacagc ttatgatcc tcgcaaacct 3420
 gagctcgact catccaaga agagctggac aagtacttca aaatcatac atccacagat 3480
 gttagctctg ggcacatttc aggcattaac gttctgtgct tcaaatctca aaagaaatt 3540
 gaccgctcca atggagtgcc taaaaattta atggaatcac tcaatgacct tcaagaattg 3600
 ggaanaatag agcaatatat taaatggcct tgg 3633

<210> SEQ ID NO 8

<211> LENGTH: 1211

<212> TYPE: PRN

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 8

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 1 5 10 15
 Ala Val Phe Val Ser Pro Ser Ala Arg Gly Ser Gly Ser Asp Leu Asp
 20 25 30
 Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln His
 35 40 45
 Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg Ser
 50 55 60
 Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser Asn
 65 70 75 80
 Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val Ile
 85 90 95
 Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn Val
 100 105 110
 Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln Ser
 115 120 125
 Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys Asn

-continued

130	135	140
Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met Gly		
145	150	155 160
Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr Phe		
	165	170 175
Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser Gly		
	180	185 190
Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly Phe		
	195	200 205
Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp Leu		
	210	215 220
Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu Gly		
	225	230 235 240
Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro Ala		
	245	250 255
Gln Asp Ile Trp Gly Thr Ser Ala Ala Tyr Phe Val Gly Tyr Leu		
	260	265 270
Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile Thr		
	275	280 285
Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys Ser		
	290	295 300
Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn Phe		
	305	310 315 320
Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr Asn		
	325	330 335
Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser Val		
	340	345 350
Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr Ser		
	355	360 365
Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly Val		
	370	375 380
Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala Asp		
	385	390 395 400
Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Gly Gln		
	405	410 415
Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Met		
	420	425 430
Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser Thr		
	435	440 445
Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu Arg		
	450	455 460
Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly Lys		
	465	470 475 480
Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp Tyr		
	485	490 495
Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val Val		
	500	505 510
Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly Pro		
	515	520 525
Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn Phe		
	530	535 540

-continued

Asn Gly Leu Thr Gly Thr	Gly Val Leu Thr Pro Ser Ser Lys Arg Phe
545	550 555 560
Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp Ser	
	565 570 575
Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys Ser	
	580 585 590
Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser Glu	
	595 600 605
Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr Ala	
	610 615 620
Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr Gly	
	625 630 635 640
Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu His	
	645 650 655
Val Asp Thr Ser Tyr Gln Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys	
	660 665 670
Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys Ser	
	675 680 685
Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala Tyr	
	690 695 700
Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile Thr	
	705 710 715 720
Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys Asn	
	725 730 735
Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Gln	
	740 745 750
Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile Ala	
	755 760 765
Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys Gln	
	770 775 780
Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe Ser	
	785 790 795 800
Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile Glu	
	805 810 815
Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met Lys	
	820 825 830
Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile Cys	
	835 840 845
Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp	
	850 855 860
Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala Thr	
	865 870 875 880
Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala	
	885 890 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val	
	900 905 910
Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala Ile	
	915 920 925
Ser Gln Ile Gln Glu Ser Leu Thr Thr Ser Thr Ala Leu Gly Lys	
	930 935 940

-continued

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
 945 950 955 960
 Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
 965 970 975
 Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp Arg
 980 985 990
 Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln
 995 1000 1005
 Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala
 1010 1015 1020
 Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp
 1025 1030 1035
 Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala
 1040 1045 1050
 Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln
 1055 1060 1065
 Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys
 1070 1075 1080
 Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser
 1085 1090 1095
 Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr
 1100 1105 1110
 Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly
 1115 1120 1125
 Ile Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp
 1130 1135 1140
 Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser
 1145 1150 1155
 Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val
 1160 1165 1170
 Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys
 1175 1180 1185
 Asn Leu Asn Gln Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
 1190 1195 1200
 Glu Gln Tyr Ile Lys Trp Pro Trp
 1205 1210

<210> SEQ ID NO 9

<211> LENGTH: 2093

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 9

atggatgcac tgaagagagg gctctgctgt gtgctgctgc tgtgtggagc agtcttcgtt 60
 tcgcccagcg ctgagaggatc ggaagtgac attgaacggt gaacaaattt tgatgatgtt 120
 caagctocta attacactca acataactca tctatgaggg gggtttacta tctgatgaa 180
 atttttagat cagcaactct ttatttaact caggatttat tcttcacatt ttattccta 240
 gttacagggg ttcatactat taatcaacg tttggcaacc ctgtcatacc ttttaaggat 300
 ggattattat ttgtcgccac agagaataca aatgtgtgac gtgtgtgggt tttgtgtctt 360
 acaatgaaca acaagtcaca gtggtgtatt attattaaca attctactaa tgttgttata 420

-continued

cgcgcctgta actttgaatt gtgtgcacac cctttctttg ctgtttcttaa acccaatgggt 480
 acacagacac atactatgat attogataat gcatttaatt gacotttoga gtacatatot 540
 gatgcotttt cgtttgatgt ttccagaaag tccaggaatt ttaaacactt acgagagttt 600
 gtgtttaaaa ataaagatgg gttttotatat gtttaaaagg gctatcaacc tatagatgta 660
 gttcgtgac tacccttcgg ttttaaacat ttgaacaacta ttttaagtt gctottttgg 720
 attaacatta caaattttag agcaattott acagcctttt cactgtctca agacatttgg 780
 ggcaogtoag ctgcgcgcta ttttgttggc tttttaaagc caactacact tatgtctaa 840
 tatgatgaaa atgtgtacat cccagatgct gttgattgtt ctcaaatcc actgtctgaa 900
 ctcaaatgct ctgttaagag ctttgagett gcaaaaggaa tttaacagac ctctaatttc 960
 agggttgttc cctcaggaga tgttgtgaga ttccctaata ttcaaacact gtgtctttt 1020
 ggagagggtt ttaatgtac taaattccct tctgtctatg cctgggagag aaaaaaatt 1080
 totaattgtg ttgtgatta ctctgtctgc tacaactcaa catttttttc aactttaag 1140
 tctatgtgct tttctgccc taagtgaat gatctttgt tctcaaatg ctatgcagat 1200
 tctttttag tcaaggagga tgatgtaga caaatagcgc caggacaaa cgtgtttatt 1260
 gctgattata attataaatt gcaagatgat ttcaatgggt gtgtccttgc ttggaact 1320
 aggaacattg atgtacttc aactggtaat tataattata aatataggta tottagaat 1380
 ggcaagotta ggccttttga gagagacata ctcaatgtgc ctttctcccc tgatggcaaa 1440
 ccttgcaccc cactgtctct taattgttat tggcattana atgattatg tttttacac 1500
 actactggca ttgctacaa accttaaaga gttgtagtac tttctttga acttttaatt 1560
 gcaacgggca cggtttttgg accaaatta tcaactgacc ttattaagaa ccaatgtgtc 1620
 aattttaatt ttaattgact cactgtact gttgtgttaa ctctctctc aagagattt 1680
 caacatttc acaatttgg cctgtatgtt tctgatttca ctgactcgt tgcagatct 1740
 aaaaactctg aaattatga cattcaact tgcctctttg ggggtgtaag tgtaattaca 1800
 cctggaacaa atgttctact tgaattgtgt gttctatata aagatgttaa ctgcaatgat 1860
 gttctacag caattcatgc agatcaact acacacgact ggagcatata ttctactgga 1920
 acaatgtat tcaagactca agcaggtgtt ctataggag ctgagcatgt cgaacttot 1980
 tatgagtgcg acattctat ttgagctggo atttgtgcta gttacactac agtttotta 2040
 tcactagta ctgacaaaa atctattgtt gttatacta tgcctttagg tgc 2093

<210> SEQ ID NO 10

<211> LENGTH: 698

<212> TYPE: PRT

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 10

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 1 5 10 15
 Ala Val Phe Val Ser Pro Ser Ala Arg Gly Ser Gly Ser Asp Leu Asp
 20 25 30
 Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln His
 35 40 45
 Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg Ser
 50 55 60

-continued

Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser Asn	
65	80
Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val Ile	
85	95
Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn Val	
100	110
Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln Ser	
115	125
Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys Asn	
130	140
Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met Gly	
145	160
Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr Phe	
165	175
Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser Gly	
180	190
Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly Phe	
195	205
Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp Leu	
210	220
Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu Gly	
225	240
Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro Ala	
245	255
Gln Asp Ile Trp Gly Thr Ser Ala Ala Tyr Phe Val Gly Tyr Leu	
260	270
Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile Thr	
275	285
Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys Ser	
290	300
Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn Phe	
305	320
Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr Asn	
325	335
Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser Val	
340	350
Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr Ser	
355	365
Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly Val	
370	380
Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala Asp	
385	400
Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly Gln	
405	415
Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Met	
420	430
Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser Thr	
435	445
Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu Arg	
450	460
Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly Lys	

-continued

465	470	475	480
Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp Tyr			
	485	490	495
Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val Val			
	500	505	510
Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly Pro			
	515	520	525
Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn Phe			
	530	535	540
Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg Phe			
	545	550	555
Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp Ser			
	565	570	575
Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys Ser			
	580	585	590
Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser Glu			
	595	600	605
Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr Ala			
	610	615	620
Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr Gly			
	625	630	635
Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu His			
	645	650	655
Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys			
	660	665	670
Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys Ser			
	675	680	685
Ile Val Ala Tyr Thr Met Ser Leu Gly Ala			
	690	695	

<210> SEQ ID NO 11
 <211> LENGTH: 1623
 <212> TYPE: DNA
 <213> ORGANISM: SARS-CoV Urbani strain
 <400> SEQUENCE: 11

atggatgcaa tgaagagagg gctctgtgct gtgctgtgct tgtgtggagc agtcctcgtt	60
tgccccagcg cttagagagac gggagatagt tcaattgctt acctcaataa caccattgct	120
atacctaata acttttcaat tagcattact acagaagtaa tgcctgttct tatggctaaa	180
acctccgtag attgtaatat gtacatctgc gggagattcta ctgaatgtgc taatttgctt	240
ctccaatatg gtatgctttg cacacaacta aatcgtgcac tctcaggtat tgcctgtgaa	300
caggatcgca acacacgtga agtgttcgct caagtcaaac aaatgtacaa aaccccact	360
ttgaastatt ttgttggttt taatttttca caaatattac ctgacctctc aaagccact	420
aaagagcttt ttattgaga ctgtctcttt aataaggtga cactcgtgta tgcctgcttc	480
atgaagcaat atggcgaatg cctaggtgat attaatgcta gagatctcat ttgtgcgag	540
aagttcaatg gacttaacgt gtggccacct ctgctcaatg atgatgatgt tgcctgctac	600
actgtgctc tagtttagtg taactgcaact gctggatgga catttggtgc tggcgctgct	660
cttcaaatac ctttgtctat goaaatggca tataggttca atggcattgg agttaacca	720

-continued

aattgtctct atgagaacca aaaaacaac gccaaccaat ttaacagcgc gattagctaa	780
attcaagaat cacttacaac aacatcaact gcattggcca agctgcaaga cgttgttaac	840
cagaatgctc aagcattaaa cacacttggt aaacaacta gctctaatt tggtgcaatt	900
tcaagtgtgc taantgatat cotttcgcga ctgtataaag tcgaggcgga ggtacaaatt	960
gacagggttaa ttacaggcag atttcaagc cttcaaacct atgtacacaa acaactaatc	1020
agggctgctg aaatcagggc ttctgctaatt ctgtctgcta ctaaaatgic tgagtgtgtt	1080
cttgaccaat caaaaagagt tgacttttgt ggaagggtgc accaacttat gtcttcoca	1140
caagcagccc cgaatggtgt tgtcttcta catgtcaact atgtgccatc ccaggagagg	1200
aacttcacca cagcgccagc aatttgtcat gaaggcaaat catacttccc tcgtgaaggt	1260
gtttttgtgt ttaatggcac ttcttggttt attacacaga ggaactttct ttctccaca	1320
taataactaa cagacatac atttgtctca ggaattgtg atgtcgttat tggcatcatt	1380
aacacacag ttatgatcc totgaacct gagctcgact cattcaaga agagctggac	1440
agtaacttca aaatcatatc ataccagat gttagctttg ggcacatttc aggcattaac	1500
gcttctgtcg tcaacattca aaaaagaatt gacgcctcca atgaggtcgc taanaattta	1560
aatgaatcac tcattgacct tcaagaattg ggaatatatg agcaatatat taantggcct	1620
tgg	1623

<210> SEQ ID NO 12

<211> LENGTH: 541

<212> TYPE: PRN

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 12

Met Asp Ala Met Lys Arg Gly Leu Cys Val Leu Leu Cys Gly	
1 5 10 15	
Ala Val Phe Val Ser Pro Ser Ala Arg Gly Ser Gly Asp Ser Ser Ile	
20 25 30	
Ala Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser	
35 40 45	
Ile Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp	
50 55 60	
Cys Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu	
65 70 75 80	
Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly	
85 90 95	
Ile Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val	
100 105 110	
Lys Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn	
115 120 125	
Phe Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe	
130 135 140	
Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe	
145 150 155 160	
Met Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu	
165 170 175	
Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu	
180 185 190	

-continued

Thr Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr
 195 200 205
 Ala Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Leu Gln Ile Pro
 210 215 220
 Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln
 225 230 235 240
 Asn Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys
 245 250 255
 Ala Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu
 260 265 270
 Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr
 275 280 285
 Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu
 290 295 300
 Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile
 305 310 315 320
 Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr
 325 330 335
 Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala
 340 345 350
 Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp
 355 360 365
 Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala Pro
 370 375 380
 His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln Glu Arg
 385 390 395 400
 Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys Ala Tyr Phe
 405 410 415
 Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser Trp Phe Ile Thr
 420 425 430
 Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr Thr Asp Asn Thr Phe
 435 440 445
 Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile Ile Asn Asn Thr Val
 450 455 460
 Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp
 465 470 475 480
 Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile
 485 490 495
 Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg
 500 505 510
 Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln
 515 520 525
 Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp
 530 535 540

<210> SEQ ID NO 13

<211> LENGTH: 1269

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 13

atgtctgata atggaccocca atcaaaccaa cgtagtgcoc coogcattac atttggtgga

60

-continued

```

ccccagatt caactgacac taaccgaat ggaggacgca atggggcaag gccaaacag 120
cgccagcccc aaggtttacc caataatact gcgtcttggt tcacagctct caactagcat 180
ggcaaggagg aacttagatt cctcgaggg caggcggttc caatcaacac caatagtgg 240
ccagatgacc aaattggcta ctacogaaga gtaacccgc gagttcgtgg tggtagcggc 300
aaatgaag agtcagccc cagatggta tctattacc taggaactgg cccagaagct 360
tcacttccct acggcgctaa caaagaaggc atcgtatggg ttgcaactga gggagccttg 420
aatacaccca aagaccacat tggcacccgc aatctaata acaatgctgc cacgtgcta 480
caacttcctc aaggaacacac attgccaaaa gcttctatc cagagggaag cagaggcggc 540
agtcagcct cttctcgtc ctcatcagct agtcgggta attcaagaaa tcaactcct 600
ggcagcagta ggggaattc tctgtctga atggtagcg gagtggtga aactgcctc 660
gcgtattg cgtagacag attgaaccag cttgagagca asgtttctgg taagggccaa 720
caacaacag gccaaactgt caataagaaa tctgtctgt aggcctctaa aaagcctgc 780
caaaaacgta ctgccacaaa acagtacaac gtactcaag catttgagg acgtgtcca 840
gaacaaccc aggaattt cggggaccaa gactaatca gacaaggaac tgattacaa 900
cattggcgc aaattgcaca attgttcca agtgcctct cattcttgg aatgtcacg 960
attggcatgg aagtacacc ttgggaaca tgggtgact atcatggagc cattaaattg 1020
gatgacaag atccacaatt caagacaac gtactactg tgaacaagca cattgagcca 1080
tacaaaacat tcccaccac agagcctaaa aaggacaaa aaaaaagac tgaagaagct 1140
cagcctttgc cgcagagaca aaagaacag cccactgtga cttctctcc tggcgtgac 1200
atggatgatt tctccagaca acttcaaat tccatgagtg gagcttctgc tgaaccaact 1260
caggcataa 1269

```

<210> SEQ ID NO 14

<211> LENGTH: 422

<212> TYPE: PRK

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 14

```

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro Arg Ile
1      5      10      15
Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn Gly Gly
20     25     30
Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn
35     40     45
Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Glu
50     55     60
Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Gly
65     70     75     80
Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Arg Ala Thr Arg Arg Val Arg
85     90     95
Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr Phe Tyr
100    105    110
Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala Asn Lys
115    120    125
Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr Pro Lys
130    135    140

```

-continued

Asp His Ile Gly Thr Arg Asn Pro Asn Asn Ala Ala Thr Val Leu
 145 150 155 160
 Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly
 165 170 175
 Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Arg Ser Arg
 180 185 190
 Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn Ser Pro
 195 200 205
 Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu Leu
 210 215 220
 Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys Gly Gln
 225 230 235 240
 Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu Ala Ser
 245 250 255
 Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn Val Thr
 260 265 270
 Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly
 275 280 285
 Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln
 290 295 300
 Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met Ser Arg
 305 310 315 320
 Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr His Gly
 325 330 335
 Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn Val Ile
 340 345 350
 Leu Leu Asn Lys His Ile Asp Ala Tyr Lys Thr Phe Pro Pro Thr Glu
 355 360 365
 Pro Lys Lys Asp Lys Lys Lys Lys Thr Asp Glu Ala Gln Pro Leu Pro
 370 375 380
 Gln Arg Gln Lys Lys Gln Pro Thr Val Thr Leu Leu Pro Ala Ala Asp
 385 390 395 400
 Met Asp Asp Phe Ser Arg Gln Leu Gln Asn Ser Met Ser Gly Ala Ser
 405 410 415
 Ala Asp Ser Thr Gln Ala
 420

<210> SEQ ID NO 15

<211> LENGTH: 1209

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 15

atgtctgata atggaccocaa atcaaaccaa cgtagtcccc cccgcattac atttggtaga 60
 cccacagatt caactgacaa taacacagat ggaggacgca atggggcaag gccaaacag 120
 cgcgcacccc aaggtttacc caataactac ggttgttggt tcacagctct cactcagcat 180
 ggcacaggag aacttagatt cctcagagcc caggcgcttc caatcaacac caatagtgtt 240
 ccagatgccc aaattggcta ctaccgaaga gctaccggac gagtctgtgg tggtagcggc 300
 aaatgaagag agctcagccc cagatggtac ttctattacc taggaactgg cccagaagct 360
 tcacttcctc acggcgctaa caaagaagcg atcgtatggg ttgcacactga gggagccttg 420

-continued

aatacaccca aagaccacat tggaccgcg aatcctaata acaatgctgc caccgtgcta 480
caacttcttc aaggaacaa attgcacaaa ggcttctacg cagagggagc cagagcgccg 540
agtcaggctt ctctctgctc ctctacactg agtcgggta attcaagaaa ttaactctct 600
ggcagcagta ggggaatttc tctgtctgca atgctgtagc gaggtgtgta aactgccttc 660
ggcgtatttg tgcagacag attgaaccgc ctgagagcga aagtctctgg taagagccaa 720
caacaaacaag gccaaactgt cactaagaaa tctgtgctg aggcattctaa aagcctcgc 780
caaaaacgta ctgcacaaa acagtacaac gtaactcaag catttgggag acgtgtgcca 840
gaacaaaccc aaggaatttt cggggaacaa gacctaatca gacaaggaac tgattacaaa 900
cattggccgc aaattgcaca atttgtacca agtgcctctg cattctttgg aatgtcaagc 960
attggcatgg aagtcacaco ttcgggaaca tggctgactt atcatggagc cattaaattg 1020
gatgcacaaag atccacaatt caaagacaaac gtcactactg tgaacaagca cattgagcca 1080
taccctttgc cgcagagaca aaagaagcgc cccactgtga ctctctctcc tggcgctgac 1140
atggatgatt tctccagaca acttcaaaat tccatgagtg gagctctctg tgattcaact 1200
caggcataa 1260

<210> SEQ ID NO: 16

<211> LENGTH: 402

<212> TYPE: PRT

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 16

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro Arg Ile
1 5 10 15

Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn Gly Gly
20 25 30

Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn
35 40 45

Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Glu
50 55 60

Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Gly
65 70 75 80

Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Ala Thr Arg Arg Val Arg
85 90 95

Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr Phe Tyr
100 105 110

Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala Asn Lys
115 120 125

Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr Pro Lys
130 135 140

Asp His Ile Gly Thr Arg Asn Pro Asn Asn Asn Ala Ala Thr Val Leu
145 150 155 160

Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly
165 170 175

Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg Ser Arg
180 185 190

Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn Ser Pro
195 200 205

-continued

Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu Leu Leu
210 215 220

Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys Gly Gln
225 230 235 240

Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu Ala Ser
245 250 255

Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn Val Thr
260 265 270

Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly
275 280 285

Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln
290 295 300

Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met Ser Arg
305 310 315 320

Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr His Gly
325 330 335

Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn Val Ile
340 345 350

Leu Leu Asn Lys His Ile Asp Ala Tyr Pro Leu Pro Gln Arg Gln Lys
355 360 365

Lys Gln Pro Thr Val Thr Leu Leu Pro Ala Ala Asp Met Asp Asp Phe
370 375 380

Ser Arg Gln Leu Gln Asn Ser Met Ser Gly Ala Ser Ala Asp Ser Thr
385 390 395 400

Gln Ala

<210> SEQ ID NO 17
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: SARS-CoV Urban1 strain
<400> SEQUENCE: 17

Lys Thr Phe Pro Pro Thr Glu Pro Lys Lys Asp Lys Lys Lys Thr
1 5 10 15

Asp Glu Ala Gln
20

<210> SEQ ID NO 18
<211> LENGTH: 666
<212> TYPE: DNA
<213> ORGANISM: SARS-CoV Urban1 strain
<400> SEQUENCE: 18

atggcagaca acggtactat tacctgtgag gagottaaac aactcctgga acastggaac 60
ctagtaastag gtttccattt cctagcctggt attatgttac tacacatttg cttattctaat 120
cgagacagggt tttgtacat aataaagctt gtttccctct ggctcttggt gccagtaaca 180
cttgcttggt ttgtgcttgc tggctgtcac agaattaatt gggtagctgg cgggattgcy 240
attgcaatgg cttgtattgt aggtctgatg tggcttagct acttctgtgc ttcttcacgg 300
ctgtttgtgc gtaccgcgtc aatgtggtca ttcaaccagg aaacaacat tctttcact 360
gtgcctctcc gggggacaat tgtgaccaga ccgtctcatg aaagtgaact tgtcaatggt 420
gctgtgatca ttgtgtgcca cttgcgaatg gccggacacc ccttagggcg ctgtgacatt 480

-continued

```

aaggacctgc caaagagat cactgtgct acatcacgaa cgttttetta ttacaaatta 540
ggagcgtgc agcgtgtag cactgatcca ggtttgctg catcacacgc ctacgtatt 600
ggaaactata aattaaatac agacacagcc ggtagcaacy acaatattgc ttgtagta 660
cagtaa 666

```

```

<210> SEQ ID NO 19
<211> LENGTH: 221
<212> TYPE: PRT
<213> ORGANISM: SARS-CoV Urbani strain
<400> SEQUENCE: 19

```

```

Met Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln Leu Leu
1      5      10      15
Glu Gln Trp Asn Leu Val Ile Gly Phe Leu Phe Leu Ala Trp Ile Met
20     25     30
Leu Leu Gln Phe Ala Tyr Ser Asn Arg Asn Arg Phe Leu Tyr Ile Ile
35     40     45
Lys Leu Val Phe Leu Trp Leu Leu Trp Pro Val Thr Leu Ala Cys Phe
50     55     60
Val Leu Ala Ala Val Tyr Arg Ile Asn Trp Val Thr Gly Gly Ile Ala
65     70     75     80
Ile Ala Met Ala Cys Ile Val Gly Leu Met Trp Leu Ser Tyr Phe Val
85     90     95
Ala Ser Phe Arg Leu Phe Ala Arg Thr Arg Ser Met Trp Ser Phe Asn
100    105    110
Pro Glu Thr Asn Ile Leu Leu Asn Val Pro Leu Arg Gly Thr Ile Val
115    120    125
Thr Arg Pro Leu Met Glu Ser Glu Leu Val Ile Gly Ala Val Ile Ile
130    135    140
Arg Gly His Leu Arg Met Ala Gly His Pro Leu Gly Arg Cys Asp Ile
145    150    155    160
Lys Asp Leu Pro Lys Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser
165    170    175
Tyr Tyr Lys Leu Gly Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe
180    185    190
Ala Ala Tyr Asn Arg Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp
195    200    205
His Ala Gly Ser Asn Asp Asn Ile Ala Leu Leu Val Gln
210    215    220

```

```

<210> SEQ ID NO 20
<211> LENGTH: 231
<212> TYPE: DNA
<213> ORGANISM: SARS-CoV Urbani strain
<400> SEQUENCE: 20

```

```

atgtactaat tcgttttcga agaacacagt acgttaaatg ttaatagcgt actcttttt 60
cttgcttttg tggattattc gctagtacaa ctacgcatcc ttactgcgtc tcgattgtgt 120
gcgtactgct gcaatattgt taacgtgagt tttagtaaac caacggttta cgtctactcg 180
cgtgttaaaa stctgaactc ttctgaagga gttcctgata ttctgtgcta a 231

```

```

<210> SEQ ID NO 21

```

-continued

<211> LENGTH: 76

<212> TYPE: PRY

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 21

```

Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu Ile Val Asn Ser
 1           5           10           15
Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala
 20           25           30
Ile Leu Thr Ala Leu Arg Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn
 35           40           45
Val Ser Leu Val Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn
 50           55           60
Leu Asn Ser Ser Glu Gly Val Pro Asp Leu Leu Val
 65           70           75

```

<210> SEQ ID NO 22

<211> LENGTH: 3768

<212> TYPE: DNA

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 22

```

atgtttattt tottattatt tottactct actagtggta gtgacctga ccggtgcacc 60
acctttgatg atgttaacgc tctaatatc actaacata ctctactat gaggggggtt 120
taactctctg atgaatttt tagatcagac actcttatt taactcagg tttattcttt 180
ccattttatt ctaagtgtac agggtttcat actattaatc atcgttttg caaccctgtc 240
atacctttta aggatgggtat ttattttgct gccacagaga aatcaaatgt tgcctgtgtt 300
tgggtttttg gttctacct gaaacaacag tcaacgtctg tgattattat taacaattct 360
actaatgttg ttatacagac atgttaacttt gaattgtgtg acaacccttt ottgtgtgtt 420
tctaaaccca tgggtacaca gacacatact atgatattcg ataagcatt taattgcact 480
ttcagataca tatctgatgc cttttcgctt gatgtttcac aaaaagtcag taattttaaa 540
cacttacag agttttgttt taataataaa gatgggtttc totatgttta taagggttat 600
caacotatag atgtagtctg tgatctacct totgttttta acatttgaa acctattttt 660
aagttgcttc ttgttattaa cattacaact tttagagcca ttctacagc cttttcaact 720
gtcacaagaa ttgggggcac gtcagctgca gccatttttg ttggtatttt aaagccaact 780
acatttatgc tcaagatgaa tgaatatggt acaactacag atgctgttga ttgttctcaa 840
aatccacttg ctgaactcaa atgctctggt aagagctttg agattgacaa aggaatttac 900
cagacotcta atttcagggg ttgtcctcaa gtagatgttg tgaattctcc taattattca 960
aacttgtctc cttttggaga ggtttttaat gctactaaat tccctctgt ctatgatggt 1020
gagagaaaaa aaattttctaa ttgtgtgtct gatctactgt tgcctacaaa ctcaactatt 1080
ttttcaacct ttaagtgtca tggcgtttct gccactaaat tgaatgatct ttgtctctcc 1140
aatgtctatg cagattcttt tttagtcaag gtagatgatg taagacaact agcgcaggga 1200
caaacctggt ttattgtcga ttataattat aaattgcac atgatttcat gggttgtgtc 1260
cttgcttggg atactaggaa cattgatgct acttcaactg gtaattataa ttataaatat 1320
aggtatctta gacatggcaa gcttagggcc tttagagag acatatctaa tgcgtcttcc 1380
tccctgtatg gcaaaccttg caacccact gctttaatt gttattggcc attaatgat 1440

```

-continued

tatgggtttt acaccactac tggcattggc taccacactt acagagttgt agtactttct	1500
tttgaacttt taactgcaac ggcacagggt tgtggaccac aattatocac tgaccttatt	1560
aagacacagt gtgtcaattt taattttaet ggactcaacty gtactggtgt gtttaactct	1620
tcttcaasga gatttcaacc atttcaacaa ttggccgctg atgtttctga tttaactgat	1680
tccgttcgag atctcaaaac atctgaanta ttgacacttt caccttgctc ttttgggggt	1740
gtaagtgtaa ttacaactgg aacaaatgct tcatatgaag ttgctgtctt atatacagat	1800
gttaactgca ctgatgttcc tacagcaatt cctgcagatc aactcacacc agcttggcgc	1860
atataattca ctggaacaaa tgtattccag actcaagcag gctgtcttat aggagctgag	1920
cctgtcgaca cttcttatga gtgcgacatt cctattggag ctggcatttg tgcagttaac	1980
catacagttt ctttattacg tagtactagc caaaatctta ttgtggttta tactatgtct	2040
ttagggtgct atagttcaat tgcattactc aataacacaa ttgotatacc taactaacttt	2100
tcaattagca ttaactacga agtaactgct gttctatggt ctaaaactcc cgtagattgt	2160
aatatgtaca tctgcggaga ttctactgaa tgtgctaatt tgcctctcca atatggtgac	2220
ttttgcaaac aactaatcgt tgcactctca ggtattgctg ctgaacagga tcgcaacaca	2280
cgtgaagtgt toytccaagt caaacaaatg tacaacaccc caactttgaa atattttggt	2340
ggttttaatt tttaacaaat attaacgac cctctaaagg caactaagag gtcttttatt	2400
gaggacttgc tctttaataa ggtgacactc gctgatgctg gcttaatgaa gcaatatgac	2460
gaatgcttag gtgatattaa tgcctagatc ctcaattgtg cgcagaagtt caatggactt	2520
acagtttgcg cacctctgct cactgatgat atgatgctg cctcaactgc tgcctagttt	2580
agtgttactg ccaactgctg atggacattt ggtgctggcg ctgctcttca aatacctttt	2640
gctatgcaca tggcatatag gttcaatggc attggagtta cccaatatgt tctctatgag	2700
aaccaaaacc aactcgcaca ccaatttaac aaggcgatca gtcaaatcca agaatcactt	2760
acacacacac caactgcatt ggcgaagctg caagacgttg ttaaccagaa tgcctaaaga	2820
ttaaacacac ttgttaacaa ccttagctct aattttgggt caatttcaag tctgtcaaat	2880
gatatccttt cgcgacttga taagctgag cggagagttac aaattgacag gtttaattca	2940
ggcagacttc aaagccttca aacctatgta acacacacac taactcgggc tgcctgaatc	3000
agggcttctg ctactcttgc tgcactaata atgtctgagt gtgtctcttg acaatcaaaa	3060
agagttgact ttgttgaaaa gggctacacc ctattgtctt tccacaagc agccccgcct	3120
ggtgttctct tctacatgct cactgatgtg ccaatccagg agaggaactt caccacagcg	3180
ccagcaattt gtcattgaag caaagcctac tctcctctg aggtgttttt ttgttttaatt	3240
ggcacttctt ggtttattac acagaggaac ttcttttctc cacaataat tactacagac	3300
aatacatttg tctcagaaaa ttgtgatgct gttattggca tcatatacaa cactgattat	3360
gatcctctgc aacctgagct cgaactcctc aaagaagagc tggacaagta cttcaaaaa	3420
catacatcac cagatgttga tcttgccgac atttcaggca ttaacgcttc tctcgtcaac	3480
attcaaaagc aattgacgc cctcaatgag gtgcgttaaa atttaaatga atcaactcatt	3540
gacctcaagc aattgggaaa atatgagcaa tatattaact ggccttggtt tgtttggctc	3600
ggcttcaatt ctggactaat tgcactctc atggttacaa tctgtgttg ttgctgact	3660
agttgttgca gttgctccaa ggggtcctgc tctgtgtgtt ctgtctgcaa gttgtgtgag	3720

-continued

gatgactotg agcagcttct caagggtgtc aaattacatt aacataa

3768

<210> SEQ ID NO 23

<211> LENGTH: 1255

<212> TYPE: PRT

<213> ORGANISM: SARS-CoV Urbani strain

<400> SEQUENCE: 23

```

Met Phe Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
 1           5           10          15

Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
 20          25          30

His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
 35          40          45

Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
 50          55          60

Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
 65          70          75          80

Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
 85          90          95

Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
100         105         110

Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
115         120         125

Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
130         135         140

Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
145         150         155         160

Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
165         170         175

Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly
180         185         190

Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
195         200         205

Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu
210         215         220

Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro
225         230         235         240

Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Tyr Phe Val Gly Tyr
245         250         255

Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile
260         265         270

Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys
275         280         285

Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn
290         295         300

Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr
305         310         315         320

Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser
325         330         335

Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr
340         345         350

```

-continued

Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly	355	360	365
Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala	370	375	380
Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly	385	390	395
Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe	405	410	415
Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser	420	425	430
Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu	435	440	445
Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly	450	455	460
Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp	465	470	475
Tyr Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val	485	490	495
Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly	500	505	510
Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn	515	520	525
Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg	530	535	540
Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp	545	550	555
Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys	565	570	575
Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser	580	585	590
Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr	595	600	605
Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr	610	615	620
Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu	625	630	635
His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile	645	650	655
Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys	660	665	670
Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala	675	680	685
Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile	690	695	700
Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys	705	710	715
Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu	725	730	735
Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile	740	745	750

-continued

Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys	755	760	765
Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe	770	775	780
Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile	785	790	800
Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met	805	810	815
Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile	820	825	830
Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr	835	840	845
Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala	850	855	860
Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe	865	870	875
Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn	885	890	895
Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala	900	905	910
Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly	915	920	925
Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu	930	935	940
Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn	945	950	955
Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp	965	970	975
Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln	980	985	990
Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala	995	1000	1005
Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp	1010	1015	1020
Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala	1025	1030	1035
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln	1040	1045	1050
Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys	1055	1060	1065
Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser	1070	1075	1080
Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr	1085	1090	1095
Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly	1100	1105	1110
Ile Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp	1115	1120	1125
Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser	1130	1135	1140
Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val			

-continued

1145	1150	1155
Val Asn Ile Gln Lys Glu Ile	Asp Arg Leu Asn Glu	Val Ala Lys
1160	1165	1170
Asn Leu Asn Glu Ser Leu Ile	Asp Leu Gln Glu Leu	Gly Lys Tyr
1175	1180	1185
Glu Gln Tyr Ile Lys Trp Pro	Trp Tyr Val Trp Leu	Gly Phe Ile
1190	1195	1200
Ala Gly Leu Ile Ala Ile Val	Met Val Thr Ile Leu	Leu Cys Cys
1205	1210	1215
Met Thr Ser Cys Cys Ser Cys	Leu Lys Gly Ala Cys	Ser Cys Gly
1220	1225	1230
Ser Cys Cys Lys Phe Asp Glu	Asp Asp Ser Glu Pro	Val Leu Lys
1235	1240	1245
Gly Val Lys Leu His Tyr Thr		
1250	1255	

<210> SEQ ID NO 24

<211> LENGTH: 3588

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized soluble S protein

<400> SEQUENCE: 24

```

atgttttatct tctctctctt cctgacgctc actgacggat ccgacttaga tcggtgtacc 60
actttcgacg acgtccaggg cctcaactat actcaacata cctccagtat ggcgggggtg 120
tactatccag atgagatttt tcggagcgac actctgtact taacacagga cctgtttota 180
cctttttatt caaatgtaac cggcttccac accattaaac atacatttgg caatcccggt 240
ataccattca aagacggcat ttacttcgoc gcaacagaaa agagcaatgt tgtgaggggg 300
tgggtcttcg gctccacaat gaacaataaa tctcsgctgt tcatcatcat caataacagc 360
actaacgtgg taatcccggt ctgcgaattc gagctttgtg acaaccattt cttgcocgtg 420
tctaagccta tgggcaccca gactcaaca atgatctttg acaatgottt caactgcacc 480
tcgaataaca tatccagatg attctctttg gatgtcagtg aaaagtctgg aaactttaa 540
catctgagag agttgtcttt caaaaacaa gacgggttcc tctacgttta caagggttat 600
cagcccaattg atgtggttgc ggaactccct tcagggttta acacattgaa accaatattc 660
aaactgcoccc tgggtatcaa tattactaac ttccagacca tctgacocgc ctttccoccc 720
gcgcagacaa tatgggggaa ccgctgcgca gctattttcg tcggttatct gaagcccaat 780
acatttatgc tgaagtacga cgagaacgga accattacgc atgctgtcga tgttcacag 840
aatccactgg ctgaattgaa atgtctcgtg aagagotttg agatcgataa ggggatttac 900
cagacgtcta atttccagat ggtccctcca ggaagtgttg tttagttccc caatataca 960
aatttggccc ccttcgggtg agtgttcaat gccacaaagt tccgtctgtt ctaagotttg 1020
gagcggaaaa agataagcaa ctgtctgcgc gattacagtg tctatataa ctcgaacttt 1080
tttagcacgt tcaagtgtta cggggtgagt gctactaaac tgaatgattt atgttttagt 1140
aacgtttatg cagactcctt tgttgtaaa ggtgatgacg tcgcgcacaa tgcacactgg 1200
cagaccggag tgatgcgaga ttataactac aaacttcacg acgaatttat gggatgcgtg 1260
ctgccttgga acactgcgaa catcgacgca accagacocg ggaactataa ttacaataac 1320

```

-continued

agatcctca ggcacggcna gctgcggcct tttagcggyg atatctcaaa cgtcccaattt	1380
agcccgagc gcaagccctg tactctccc gacttaact gttactggcc actgaacgat	1440
tatggctttt ataccacaac cggcctcgcg taccagccct aocgggtggt ggtgatctct	1500
ttcgagctgc tgaacgcgcc tgcacccgta tgtgggccc agctttcgac agatctcacc	1560
aagaacccat gcgtaatatt caatttcaat ggccttacag gaacgggtgt gctgacccc	1620
tctccaaaga ggtttcaaac ttccacgacg ttggagctg acgtctcaga ctttactgac	1680
agtgtgaggy atctaaagac ctctgaatc ctggatata ctccctgttc cttcgtggg	1740
gttagtgtga taacccctgg gacaaatgct agttccgaag tggcgtact ctatcaagac	1800
gtgaactgca cagacgtgtc aacccgcac caccgtgac aactcaacc ggttggggy	1860
atctatagca ctggcaataa cgtgttccaa acgcagcccg gctgcttat aggggcaag	1920
catgtcgaca cttcttacga gtgtgatata ccaatcgagg ccggcatctg cgcctcaaac	1980
cacacggtag cttgtctgag ctccacagct caagaagata ttgtcgata caccatgca	2040
ctcgccgag attcaagtat cgcctacagc aataaacta tgcctattcc taccacatt	2100
tccatttcca tcaacactga ggttatgct gtctccatgy ctaagacttc cgtggactgc	2160
aatatgtaca ttgttgggga ctctacagag tgcgtaccc tttaactgca gtatggtccc	2220
ttctgcacac agctgaatag agccctgagc ggaattgccg ctgagcgagg tagaatacgc	2280
agagaagtgt ttgccagcgt gaacagatg tataagactc caaccttga gtattogga	2340
gggttcaatt ttacgaagat ccttctgac cccttgaaag cgcacaaaag gagcttcac	2400
gaagatcttc tgttcaacaa agttacttta cgggacgcgg gtttatgaa acagtatggc	2460
gagtgtctcg gggatattaa ttcccgcgat ctactctgtg ctcaagaatt caacggcctc	2520
acagtgtctc ccccaacttc gacggatgat atgatgcgcg cttaacacagc cgcactcgtg	2580
agcggcaccg ccacagccgg ttggacattc ggaactggag ccgcatcaca gattcattc	2640
gctatgcaga tggcgtacag gtccaacgga ataggcgtga cccagaacgt gttgtatgaa	2700
aatcagaagc agatttcgaa ccagttcaac aaagccattt ctcaaatcca ggaatccctg	2760
accacccaaa gcaaggcaat gggaaagctg caagacgtgg tcaaccagaa cgcaccaacc	2820
ctaaataccc tggttaaagca cgtgtctagc aattttgagc cgtttctatc tgtcctaac	2880
gatatactat caagactgga caaagtggag gcagaggtcc aactgcaccg cctgattacy	2940
ggccgcctcc agagccttca gactatgtg accaagcagc tgataagagc tgcgttaata	3000
cgagctcggc ctactctgag cgcacacaaa atgtccgatg gctcctggg gcagtccaaa	3060
cgtgtcgatt tctgvcgaaa aggttaccat ttgatgtcat ttccacagc ggctcctcac	3120
ggcgtagtgt tctctgcagt gacttatgta ccttcgcagg aaaggaactt cacaactgac	3180
ccagccatct gccatagggy aaagcatat ttcccocgag aaggtgtttt ogttttaaac	3240
gggacaagct ggttcattac tcaaggaaat tttttctgc cagagatcat taocactgat	3300
aacacatttg tatctgttaa ctgcgacgta gttatcggga ttatcaataa taacgtctat	3360
gaacccctgo aacctgagct ggaatgcttt aagagaagag tggacaagta ctttaagaat	3420
cacacctctc cagacgtgga cctgggagac atctcggca ttaatgcaag tgttgtgaat	3480
attcagaagag agattgatag actaaacgaa gtgtcgaaga actgaatga gagtttaatt	3540
gacctacagg agctcggtaa gtacgaacag tacatcaaat ggcctgtg	3588

-continued

```

<210> SEQ ID NO 25
<211> LENGTH: 3588
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Uniform optimization of S protein

<400> SEQUENCE: 25

atgttcattct tctgtctgtt cctgacctg accagcgca gcgacctgga cgggtgcacc 60
accttgaagc acgtggaagg ccccaactac acccagcaca ccagcagcat cggggcgctg 120
taataccccc acgagatctt ccggagcgac acctgtacc tgaccagga cctgttctctg 180
cccttctaca gaaactgac cggcttccac acctataaac acaccttcg caaccctctg 240
atccccctta aggaaggcat ctacttcgac gccacgaga agagcaacgt ggtgcggggc 300
tgggtgtctg cgagcaccat gaacaaag agcagagcg tgatcatcat caacaacagc 360
accaacgtg tgatccgggc ctgcaacttc gagctgtgcg acaaccctt cttgcgctg 420
agcaagccca tgygcacca gaccacacac atgatcttg acaagcctt caactgcacc 480
ttcgagtaca tcagcgacgc cttaagcctg gacgtgagcg agagagcgcg caacttcaag 540
caactgcggg agtctgtgtt caagaacaa gacggcttcc tgcactgtga caagggtcac 600
cagcccatcg acgtgtgtcg ggaactgcgc agcggcttca acaacctgaa gccactcttc 660
aagctgcgcc tgggcntcaa catacacaac ttcggcgcca tctgaacgc cttaagcccc 720
gcccaggaac tctggggcac cagcgccgac gctactctg tgggtacac gaagccacac 780
accttcattc tgaagtacga cgaagacgac acctacacg acgcctggg ctgacgccag 840
aaccctctgg ccgagtgaa gtgcagcgtg aagagcttc agatgacaa gggcatctac 900
cagacacgca acttcgggt ggtgccacgc ggcgagctg tgcggttccc caacatcac 960
aacctgtgac ctttcggcga ggtgttcaac gccacaaagt tcccagcgt gtacycctg 1020
gagcggaaga agatcagcaa ctgcgtgccc gactacagcg tgcgtacaa cagcaacttc 1080
ttcagcaact tcaagtgotc cggcgtgagc gccacaaagc tgaacgacct gtcgttcagc 1140
aacgtgtacg ccgacagctt cgtgtgaa ggcgaagcag tgcgcagat cgcctccggc 1200
cagacgcygc tgatgcgca ctacaactac aagctgcccg acgacttcac gggctgcgtg 1260
ctgycctgga acacccgyaa catgacgccc accagcagcg gaaactacaa ctacaagtac 1320
cgtacctgca gcaagcgcaa cgtgcggccc ttgagcggg acatcagcaa cgtgccttc 1380
agccccagcg gaaagccctg caaccccccc gccctgaact gctactggcc cctgaacagc 1440
tacggtcttc accaccaacc cgcactcggc taccagccct accgggtggt ggtgctgagc 1500
ttcagctcgc tgaagccccc cgcaccgctg tgcggcccca agctgagcac cgacctgac 1560
aagaaccagt cgtgtaactt caacttcaac ggcctgacgc gcaaccggct gctgaccccc 1620
agcagcaagc gtttcagccc cttaacagag ttggccggcg acgtgagcga cttcacccag 1680
agcgtgcggg accccaagac cagcgagtc ctgagatca gccctgcag cttcggcggc 1740
gtgagcgtga taacccccg caccacagcc agcagcgagg tgcgctgct gtacacagac 1800
gtgaactgca ccgacgtgag caccgccttc cagcgcgacc agctgacccc cgctcggcgy 1860
atctacagca ccgcaacaaa cgtgttccag accaagcgcy gctgctgat cggcgccgag 1920
ccgctggaca ccagctacga ytcgagatc cccatcggg ccgcatctg cgccagctac 1980

```

-continued

cacacccgtga	gctctgtgctg	gagcaacagc	cagaaagaga	tcgtggccta	caccatgagc	2040
ctggggcgccg	acagcagcat	cgcctacagc	ascaaacacca	tcgccatccc	caccaacttc	2100
agcatcagca	tcacccaccga	ggtgatgccc	gtgagcatgg	cgaagaccag	cgtggactgc	2160
aacatgtaca	ctctgcgcga	cagcaacagc	tgcgcacacc	tgctgtctga	gtacggcagc	2220
ttctgcaccc	agctgaaccc	ggccctgagc	ggcatcgccg	cagagcagga	cgggaacacc	2280
cgggagggtgt	tcgcccaggt	gaagcagatg	tacaagaccc	caacccctgaa	gtactctggc	2340
ggcttcaact	tcagccagat	ctgcgccgac	cccttgaagc	caacccagcg	gagcttcact	2400
gaggacactgc	tgttcaacaa	ggtgaacctg	gcgcagcccg	gcttcatgaa	gcagtacggc	2460
gagtgctctg	ggcagcatca	gcgccgggac	ctgatctgcy	ccagaagtt	caacggcctg	2520
acogtgctgc	ccccccctgt	gaccgcagac	atgatcgccg	ctacacccg	cgcctgtgtg	2580
agcgccaccc	ccaccccgcc	ctggaccttc	ggcgccggcg	cgcacctgca	gattcccttc	2640
gcactgcagca	tggcctacgc	gttcaacggc	atcgggctga	ccagaaagt	gctgtacagc	2700
aaccagaagc	agatgcgcaa	ccagtccaac	naggccatca	gccagatcca	ggagagcctg	2760
accacaccca	gcaccgccc	ggcgaagctg	caggacgtgg	tgaaccagaa	cgcacaggcc	2820
ctgaacaccc	tggtagaaga	gctgacagc	aacttcggcg	ccatcagcag	cgtgctgaac	2880
gacatcctga	gcggctgga	caagtgagag	gcccaggtgc	agatgcaccc	gctgatcacc	2940
ggccggctgc	agagcctgca	gacctacgtg	accacagcgc	tgatccgggc	cgcgcagatc	3000
cgggcccagc	ccacactggc	cgcacccaag	atgacagagt	gctgtctggg	ccagagcaag	3060
cgggtggact	ctctgcgcga	ggcctaccac	ctgatgagct	tccccaggc	cgccccccac	3120
ggcggtgtgt	tctgtacagt	gacctacgtg	ccagccagag	agcggaactt	caccaccgcc	3180
cccgccactc	gccacgaggg	caagggcctac	tccccccggg	agggcgtgtt	cgtgttcaac	3240
ggccaccagc	ggttcatcac	ccagcggaac	ttcttcagcc	ccagatcat	caccaccgac	3300
aaccacttcg	tgagcggcga	ctgcgacgtg	gtgatcggca	tcatcaacaa	caccgtgtac	3360
gacccccctgc	agcccgagct	ggacagcttc	aaggaggagc	tggacaagta	cttcaagaac	3420
cacacacagc	cgcagctgga	ctggggcgac	atcagcgcca	tcaacgccag	cgtgtgtgaac	3480
atccagaagc	agatgcacgc	gctgaacagc	gtgcacagca	acctgaacga	gagcctgntc	3540
gacctgcagc	agctggggcga	gtacgagcag	tacatcaagt	ggccctgg		3588
<210> SEQ ID NO 26						
<211> LENGTH: 2049						
<212> TYPE: DNA						
<213> ORGANISM: Artificial Sequence						
<220> FEATURE:						
<223> OTHER INFORMATION: Fully Optimized soluble S1 protein						
<400> SEQUENCE: 26						
atgtttatct	ttttctgttt	tctcaatta	acttcggggt	ctgacctgga	cgggtgcacc	60
acattcgatg	acgtccaagc	ccccaactac	actcagcata	catctagcat	gcggcgcgctg	120
tactacccag	atgagatctt	taggtccgac	accctttatc	tgaaccagga	ctttttcttt	180
ctttctact	ctaatgtaac	tgggttccat	accatcaacc	ataccttttg	caacccagtg	240
attccattta	aggatgggat	ttacttcgcc	gcgacccaga	aatacaatgt	gtgcgcggcg	300
tgggttttgc	gctccacatc	gaacaataag	agtccgtccg	taattatcat	taacaatagt	360

-continued

acaaacgttg tgatcaggtc atgtaatatt gaattgtgcy acaaaccttt ctctcgtgta 420
 agcaaaccca tggggagcga gactcacagc atgatcttcy ataacgctt caattgcacg 480
 tttagtaca taccgctgc cttttctcta gctgtgtcgy aaaaacaggy gaattttaag 540
 caactgagag agtttgttct taagaacaag gacggtttct tgtacgtgta caagggtac 600
 cagcogctcg acgtgtgtcg ggaactaccc agcggattca acaacctcaa gccacttttt 660
 aagctccca cgggtatcaa tataactaac ttacagagcca ttctcacagc ttctctcca 720
 gctcaggata ttgggggac tagtgcggca gcttatttcg tgggatacct taagccaca 780
 accttcattg tgaatacaga tgaagacgga accataacct agcaggtga ctgctcacag 840
 aacccctcg cagagttgaa atgctcagtt aaatccttgc agatgcacaa ggggtattac 900
 cagacacgta actttagagt cgtgcgctca ggcagcgtcg tgaagtttcc taacatcaca 960
 aatctatgct ctctcgagga agtgttcaat gccacaaagt tcccccaggt gtacgcctgg 1020
 gagcgaaaaa agatatctaa ctgcgtgcga gactcacagc tactgtataa cagcactttt 1080
 ttacgacct ttaagtgtta tggggtgtca gcaacaaac tgaacgatct ctgcttttca 1140
 aacgtttatg ccgattcctt cgttgtcaag ggaagcagtg tccgtcaaat tgcctccggg 1200
 caaactggcy ttatcgtgca ctataactat aaactgccag acgattttat ggggtgtgct 1260
 ctogostgga ataagcgaa catogctgag acctatacgg gaactacaa ctataaatat 1320
 aggtatcttc ggaacgggaa attacggcgg ttacagcgag atatttgaa cgtgouttcc 1380
 agtcccgatg gaacacatg tactctctca gccctcaatt gttactggcc attgaatgac 1440
 tacygggtct acacgacaa tggaaatagg tatcagcctt atcgtgtcgt cgttctttct 1500
 ttogaactgc tgaatgctcc cgcacaggtg tggggtcaa aactcagcgc agactgatc 1560
 aagaatcagt cgtgaattt caatttcac ggcctgacag gcaacggcgt tctgaccca 1620
 agctcacagc gcttcacgac ctccacgcaa ttggcaggg atgtgtcga cttacagat 1680
 tcaagtgcag atcccaagac cagtgaata ctgacattt ctccgtgtag ctttgccgac 1740
 gtgtctgca ttactcctgg gacgaatgcc toagcagag tggcgtgtt atatcagac 1800
 gtttaattga cagacgtcag tacogccata catgctgac agctgactcc tgcattgaga 1860
 atctactcca caggaataaa tgtgtttcag acacagcag gttcgtgat cggagccgaa 1920
 caogtgaca caagctacga atgtgatato cctatcggtg ccggcatctg cgtatgtat 1980
 cacaagtaa gctcgtcgcg gacacagat cagaagtcca ttgtggcta tactatgtcc 2040
 ctggggcgc 2049

<210> SEQ ID NO 27

<211> LENGTH: 2049

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Uniform optimization of soluble S1 protein

<400> SEQUENCE: 27

atgtcatct tctcgtgtt cctgaacctg accagcgcca ggcactgga cagatgcacc 60
 acctcgacy acgtgcaggo ccccaactac acccagcaca ccagcagcat gagagcggtg 120
 tactaacccg acagatcttt cagaagcgac accctgtacc tgacccagga cgtgtctctg 180
 ccttttaca gcaacgtgac cggcttcac acatcaacac caactctcgg caaccocgtg 240

-continued

atccccctca aggcaggcat ctacttgcg gccaccgaga agagcaacgt ggtgagagc 300
 tgggtgttgc gacgacacat gaacaacaag agccagagcg tgatcatcat caacaacagc 360
 accaacgttg tgatcagagc ctgcaacttc gacgtgtgcg acaacccctt ctgcgcgtg 420
 agcaagccca tgggcaccca gaccacacac atgatcttcg acaacgacct caactgcacc 480
 ttgcagtaga tcagcgagcg cttaagcctg gacgtgagcg agagagagcg caacttaag 540
 caccctgagag agtttgtgtt caagaacaaag gacggtcttc tgtacgtgta caaggtgtac 600
 cagcccatcg acgttgttag agaacctgcc agcggttcca acaacctgaa gcccatcttc 660
 aagctgcccc tgggcatcaa catcaacnac ttacagagca tctgacccgc cttaagcccc 720
 gccacagaca tctggggcac cagcgccgcc gccatcttcg tgggtacct gaagcccacc 780
 acctcatgc tgaagtacga caggaacggc acctacaccg acgccgtgga ctgacgacag 840
 aaccccttgg ccagagctgaa gtgcagctg agagcttcg agatgacaaa gggcatctac 900
 cagacacaga acctcagagt ggtgccacgc ggcgacgtg tgagattccc caacatcacc 960
 aaactgtgcc ccttcggcga ggtgttcaac gccacaaagt tccccagct gtacgcttg 1020
 gagagaaaga agatagcaaa ctgcgtgccc gactacaagc tgcgtacaaa cagcaccttc 1080
 ttacgacacat tcaagtgtca cgcgtgagc gccacaaagc tgaacgaact gtgctttagc 1140
 aacgtgtacg ccagacagctt cgtgtgaaag ggcagagacg tgagacagat agcccccggc 1200
 cagacccgcy tgatgcgcga ctcaacatca aagctgcccg acgaactcat gggctgctg 1260
 ctggcctgga acacacagaa catcgacgcc accagacccg gcaactacaa ctcaagtag 1320
 agataacctga gacacggcaa cgtgagaccc ttgcagagag acatcagcaa cgtgcccttc 1380
 agccccagcg gcaagccctg caccoccccc gccctgaact gctactggcc cctgaacagc 1440
 taaggcttct acaacacacac cgcacatggc taccagccct acagagtggg ggtgctgagc 1500
 ttgcagctgc tgaagccccc cgcacccctg tgcggcccca agctgagcac cgcactgac 1560
 aagaaccagt cgtgaactt caacttaaac ggcctgacgc gccacggcgt gctgaccccc 1620
 agcagcaaga gatctcagcc cttaacagag ttgcgagag acgtgagcga cttaacagac 1680
 agcgtgagag accccaagac cagcgagatc ctggaactaa gccctgcag cttaagcgcc 1740
 gtgagcgtga tcaaccccg caccacagcc agcagcagcg tgcacctgct gtaccagcac 1800
 gtgaactgca ccagcgtgag caccgcatc cagcgcagac agctgaaccc cgcctggaga 1860
 atctacagca ccggaacaaa cgtgttccag accaagcccg gctgcctgat cggcgccagc 1920
 cactgagaca ccagctacga gtgcagatc ccatagggcg ccggcatctg cgcagctac 1980
 caaacctgga gccctgtgag aagcaacagc cagaagagca tctgtgcta caccatgagc 2040
 ctggggcgcc 2049

<210> SEQ ID NO 28

<211> LENGTH: 1539

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized 82 protein

<400> SEQUENCE: 28

gacagttcaa tgcgtatctt gaacaacact atagcaatcc caacaattt ttcaatttct 60
 ataacaacag aggtgatgcc agtgcctatg gcaagacta gcttagacty caatatgtac 120

-continued

atctgcggag attctacaga atgtgcacac ttgtgtctac agtatggatc gttctgtacc	180
cagctcaacc gggcgctgag cggcattgct gccgaacagg atcgcaatac gagagagggt	240
tttgtctcaag tgaanaacaat gtataagacc ccacacattga aatacttcgg tggattcaat	300
ttcagtcaga ttctgcacaga cccactcaaa cccacccaaga ggagctttat tgaagatcct	360
ctgttcaaca aagttacott ggccgacgct gggtttatga agcaatacgg tgaagtgcctg	420
ggcgacatta acgcaacaga cctgatctgc gcccaagaat ttaagggct caccgtttta	480
cgcgcactgc tgaactgatg tatgattgac gcttacctg cggccttctg gagtgtacc	540
gcaactgtgt gctggagctt tggcgctgg gccgccttac agatcccttt tgcacgtcag	600
atggcctaca ggttcaatgg aattggtgtc actcagaagt tctgtacga gaaccagasa	660
cagatcgcca accagttcaa taaagctatt tcacagattc aggaactact taccacaact	720
tcacaggcac tcggtaaact gcaggacgtg gtgaatcaga acgctcaggc actaataaca	780
ctcgtcaacg aactgagttc caatttcogg gcccatctca gogatttga cgaactctc	840
agtggctcg acaagtgga ggccgaagtc caaatagacc gtcttatcac aggcagacta	900
cagtcaattg agactacgt taccacagag ttgatccgag ccgctgagat acgagcctcc	960
gcaactctgy ccctacacca aatgtctcag ttgtgtctgc gacaagtaa cggggtggat	1020
ttttgcggca agggctatca cctcatgtcc ttccctcagg cagcacccca cggagtctgt	1080
ttctctgagt tgacatacgt gctcagccag gagagaaact ttacactgc gccctgcatt	1140
tgctatgag gcaaaactta ttttccccg gagggggtgt tcyttttcaa cggaaactag	1200
tggtttatca cacaaaggaa tttctctcc cccagatcca tcacaccca caaaccttt	1260
gtctctgaa actgtgacct cgttatagga atcatcaata atacagtata cgtccctctg	1320
cagcccgac actgactott caaggaggaa ctagataagt acttoaagaa tcacccagc	1380
ccggatgtag atttagggga tattagcggg attacgcct ccgtggtcaa catccaaaa	1440
gagattgaca gactgacaga agtggcgag aacctgaatg agtccctgat cgtctctcag	1500
gagctgggca agtatgaaca gtatatcaag tggccttg	1539

<210> SEQ ID NO 29

<211> LENGTH: 1539

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Uniform Optimization of 62 protein

<400> SEQUENCE: 29

gacagcagca tcgctacag caacaacacc atcgccatcc ccaccaactc cagctacaga	60
atcaccaccg agtgtatgcc cgtgagcatg gccaaagaca gctgggactg caacatgtac	120
atctgcggcg acagcaccga gtgcgccaac ctgtctgtcg agtaaggcag cttctgcacc	180
cagctgaacc gggcccttag cggctatgco gccagacagg accggaacac ccgggagggt	240
ttcgcccgag tgaagcagat gtacaagacc cccaccttga agtacttcgg cgggttcaac	300
ttcagccaga tctgcgcga cccctgaag ccacccaagg ggagttcat cgaaggacctg	360
ctgttcaaca aggtgacct ggccgacgco ggtctcatga agcagtcag cyagtgcctg	420
ggcgacatca acgcccggga cctgatctgc gcccaagaat tcaacggcct gacctgtctg	480
ccccccctgc tgaccagaga catgatcgco gctacacgy ccgcctcgtg gagcgacacc	540

-continued

gcaaccgccc gctggaccct cgcgcgcgcg gccgcctgc agatccctt cgcctgcag	600
atggcttacc ggttcaacgg catcgcgctg acccagaacg tgcgtacga gaaccagaag	660
cagatcgcca accagttcaa caaggccatc agccagatcc agggagcct gaaccacc	720
agcaccgccc tggcnaagct gcaggacgtg gtgaaccaga acgcccagcg cctgaacac	780
ctggtgnagc agctgagcag caactctggc gccatcagca gcgtgtgaa cgacatctg	840
agccggctgg acaaggtgga ggcgcgggtg cagatcgacc ggtgatcac cggccggctg	900
cagagcctgc agacctacgt gaccacgacg ctgctcggg cgcgcgagat ccgggcacgc	960
gccaccctgg ccgcacccaa gatgagcgag tgcgtgctgg gccagcgaa gcgggtggac	1020
ttctcggyca agggctacca cctgatgagc ttccccccag ccgcccccca cggcgtggtg	1080
ttcctgcacg tgacctacgt gccaccgacg gagcggaaat tcaaccacgc ccccgccatc	1140
tgcacagagg gcaaggccca cttccccccg gagggcgtgt tcgtgttcaa cggcaccagc	1200
tggttcatca cccagcgga cttcttcagc cccaagatca tcaaccagca caaaccttc	1260
gtgagcggca actgcagcgt ggtgatcgcc atcatcaaca acacgtgtga cgaccctctg	1320
cagcccgagc tggacagctt caaggaggag ctggaacaat acttaagaa ccacaccagc	1380
ccgcagctgg acctggcgca catcagcggc atcaacgca gcgtgtgtaa catccagaag	1440
gagatcgacc ggtgaacga ggtggccaag aacctgaacg agagcctgat cgaactgcag	1500
gagctgggca agtacagca gtacatcaag tggccctgg	1539

<210> SEQ ID NO 30

<211> LENGTH: 3633

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized TPA-S protein

<400> SEQUENCE: 30

atggatgcaa tgaagcgggg cctgtgctgc gtgcctctgc tctgcggggc ggtgtttgtg	60
agccccagtg ccagaggtag cgcgcgcgat ttgatagagt gcaccacatt tgatgacgtg	120
caggctccca attacacca gccaccagat tctatgagag ggtatacta ccttgacagc	180
atcttcgcga gtgatccct attattaaca caagatttat tcttaccctt ctactccaac	240
gtcacagggg ttaccacact caaccacac ttccgcaacc ccgtgatccc gtttaagat	300
ggcatttat ttgcagccac agagaagtcg aatgtagtgc ggggttggtg gtttgatca	360
acatgataa ataatctca gtccgtgato attattaaca acttaacgaa tgtggtata	420
cgagcctgta atttagagt attcgataat ccatttttgc cgttagtaaa accaatgggc	480
actcagaccc atacgatgat ttctgataac gcattcaatt gtacgttga atacattctt	540
gatgctttt cactgcagct ttccagaaaag cctgggaact tcaagcattt aagagagttc	600
gtcttataaa ataaagacgg gtctctgtac gtgtataaag gataccagcc tatcgacgtg	660
gtgcggggacc tgccaagcgg tttaataacc ctgaagccca tctttaagct gccccctggga	720
atcaatatta caaacttcag ggtatacttc accgccttta gccacgctca ggacatatgy	780
ggaaacctcg ccgcgcgcta cttctogga tatttgaaac caaccacatt catgctgaag	840
tatgacgaaa atgggacgat taccgacgac gtgactgta gtcgaaccc tttggcggag	900
ttgaagtgtc cagtcaagag ctttgagato gacaagggaa tttatcaaac tagcaacttc	960

-continued

agggtgggtgc cctccggaga tgtagtctgc ttcoccaaca tcaccaacct gtgcccggtc	1020
ggtgaggtgt ttaatgcaac taactcccc tcaagtgtatg cctgggaag aaagaaaatt	1080
agcaactgtg ttgcogatta cagctctctt tataactcaa cattctcttc taactttaag	1140
tgtctatggtg tgtccggcac taagtgaac gacctctgct ttagtaacgt gtacgtgat	1200
tcctctgtgg tgaagggyga tgacgtgcgt cagattgcac cgggcccagac cggagtaata	1260
gcgattaca attacaagtt gcttgacgac ttoatgggtc gcttctagc atggaatacc	1320
cgcaacatag atgcacacct aacygggaac taacaactaa agtacagata tctgagacac	1380
ggttaagctg ggccttttga gcgggatate tccaatgtgc ctittagccc cgaaggcaaa	1440
ccatgcaccc cactgacct gaattgttat tggcctttga acgattatgg attctacact	1500
accaactggga tcggttatca accataccgg gtctctgtcc tgagttttga actcttgaac	1560
gcgctgcaa cagtctcggy acccaagctg tcgacagacc ttatcaagaa tcaagtgtgtg	1620
aactttaact tcaatgggtc caccgctacc ggtgttctga ctccactag taagcgattt	1680
caaccattcc aacagttcgg ccgtgacgtt tccgatttta cggattcggg cgtgatcca	1740
aaaacatcag agatccttga catatgcgg tgtctctttg gaggcgtgtc tgtgattaca	1800
ccaggcacta atgtctagtg cgaactcgtc gtactatacc aggacgtgaa ctgaacggac	1860
gtgagcaggy caatccacgc tgacacgctg acccccgcct ggcgcattcta cagtaacggc	1920
aataacgtct ttcagaccga ggcgcgtgt ctgattgggy ctgagcacgt cgacacttcc	1980
tatgaatgtg atattcccat cggcgttga atttgtgcta gctacacac agtctccctt	2040
ttaaagatca ccagccagaa atctatttgt gtttacaca tgtctctcgg cgcagactca	2100
tcaattgctt atagacaaca taaccttga atccatcaaa attttagtat atccaaacc	2160
accgaggtga tgcctgtgtc tatggcgaaa acttccgtcg attgcaacat gtatatctgc	2220
ggggactcca cagaatgcgc caacctgctt ctgcagtatg gaagcttctg taactaaact	2280
aaccgcgcct tgtctgggtt tgcgcgcgag caggatagaa atactagaga ggtgttgcct	2340
caggttaaac aaatgtacaa gacacgcaca cttaagtaact tcggagggtt taactttccc	2400
cagatactcc ctgaacctct aaagctact aaacgcagt tcatcgagga tctcctgttt	2460
aataagggtg cactgcgcga tctgtgcttc atgaanacat accgagaagt cctgggagac	2520
attaacgcga gacacctgat ctgtgccacg aagttcaacy gtctgacagt acttctccc	2580
ctctgcacgg acgacatgat tctgcatac acagccgcct tagttagcgg cacagccaca	2640
gctgggttga cctttggcgc tggcgcagcg ttgcagattc cattccogat gcagatggct	2700
taccgattta accggatcgg cgtgactcag aatgttttgt atgagaacca gaacagacatc	2760
gctaatcagt ttacaacggc aatcagccag atacaagant ctctgactac cacaaacacc	2820
gctctgggaa aactgcagga cgttgtgaat cagaatgcac aggcctcaa cactcgtgtg	2880
aagcagctta gttccaattt cggggacatc tctccggttt taatatgatat cctgagtgcg	2940
ctggacaagg tcyaggccga agttcagatc gaccgctga tcaacaggag cctgaactca	3000
ttgcagactt acgtgactca gcagctcata agggctgcag agattagggc ctctgcacac	3060
cttgccgcga ccaagatgtc cagatgtgtt ctcggtcagt ccaaacgggt tgacttttgt	3120
ggcacaagctt accatctgat gacgtccccc caggccgcac ccatgaggct agtctttgtg	3180
cactaactt atgtgccatc ccaagaaagg aacttcaata cgggcgcagc catatgcaat	3240

-continued

gaaggtaaag catatttccc	tcgagaagggt gtatttgttt	tcaacgggac tagctggttt	3300
attacgcago ggaatttctt	ctcacaccacaa atcatcaacta	ctgataaacac attctgcagc	3360
ggcaattgtg acgtctgtcat	tgsaattata aacaacactg	tgtacgatcc tctgcagccg	3420
gaactgtgatt ctttttaagga	ggagctcgac aagtacttca	aaaaccatac ctgcgccgac	3480
gtgyacctag gcgatattctc	tgyggttaat gcctcagtag	tcaacatcca gaaggagata	3540
gaaccgactta atgaggttgc	caagaatctg aatgagagtc	tcctgatctc gaaagaaact	3600
ggcaagtatg acaaatatat	caaatggcca tgg		3633

<210> SEQ ID NO 31

<211> LENGTH: 3633

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Uniform optimization of TPA-S protein

<400> SEQUENCE: 31

atggacgcga tgaagcgggg	cctgtgctgc gtgtgtgtgc	tgtgoggcgc cgtgtctgtg	60
agccccagcg cccggggcag	cggcagcgac ctggaacggt	gcacacactt cgacgacgtg	120
caggccccca actaacacca	gcacaccago agcatgoggg	gctgtactac ccccgacgag	180
atcttccyga gcgacacact	gtacctgacc caggacctgt	tcctgccttt ctacagcaac	240
gtgacccgct tcacacacct	caaacacacc ttcggcaacc	ccgtgatccc cttaaggac	300
ggcatctact tcgcgcgccc	cgagagaagc aacgtggtgc	ggggctgggt gttcggcagc	360
accatgaaca cnaagagcca	gagcgtgctc atcatcaaca	acagacccaa cgtgtgtgac	420
cgggcttgca acttcgagct	gtgcgacaac cctttcttgc	ccgtgagcaa gcccatgggc	480
acccagaccc acacccatgat	cttcgacaac gccttaactc	gcaccttcga gtacatcagc	540
gacgccttca gcctggacgt	gagcgagaaq agcggcaact	tcaagacact gcgggagttc	600
gtgttcaaga acaaggacgg	cttctctgac gtgtacaagg	gtacacagcc catcgacgtg	660
gtgcggggacc tgcaccagcg	cttcaacacc ctgaagccca	tcttcaagct gccctgggce	720
atcaacatca ccaacttcgg	ggccatctgt accgccttca	gccccgcgca ggcacatctg	780
ggcaaccagcg ccgcgcgcta	cttcgtgggc tactcgaaac	ccacacactt catcgtgaag	840
tacgacgaga acggcaccat	caccgacgac gtggaactga	gccagaaccc cctggccgag	900
ctgaagtgcg cgtgtaaagg	cttcgtgagtc gacaaaggca	tctaccagac cagcaacttc	960
cggtgtgttgc ccaagcggga	cgtggtgcgg ttcccacaac	tcacaaacct gtgccccttc	1020
ggcgaggtgt tcaacgccac	caagttccccc agcgtgtacg	cctgggagcy gaagaagata	1080
agcaactcgc tggccgacta	cagcgtgtgt tacaacagca	cttctctcag cacctcaag	1140
tgtctacggcg tgagcgccac	caagctgaac gaactgtgct	tcagcaacgt gtacgcgac	1200
agcttcgtgg tgaaggcgga	cgactgtcgg cagatgcgcc	ccggccagac cggcgtgata	1260
gccgactaca actcaagct	gcccgacgac ttcattgggt	gctgtctggt cttggaacac	1320
cggaacatcg accccaccag	cacgggcaac tacaactaca	agtaacgcta cctgcggcac	1380
ggcaagctgc gcccttcga	gcgggaacac agcaacgtgc	ccttcagccc cgacggcaag	1440
ccctgcaccc ccccccgcct	gaactgtgac ttggccctga	acgactacgg cttctaacac	1500
accaaccgca tcggtacaca	gccctaaccg gtgggtggtg	tgaacttcga gctgctgaac	1560

-continued

```

gccccgccca ccgtgtgcgg ccccaagctg agcacagacc tgatcaagaa ccagtgctg 1620
aacttcaact toacagcgctt gaccggcacc ggctgtgtga ccccaagcag caagcggttc 1680
cagcgccttc agcagctcgg ccgggagctg agcgacttca ccgacagcgt cggggacccc 1740
aagacacagc agatcctgga cctcagcccc tgcagcttcg cggcgctgag cgtgataacc 1800
cccgccacca accgccacag cgaagtggcc gtgtctgacc aggcagtga ctcgacccgc 1860
gtgagcaccg ccatccacgc cgcacagctg acccccgcct ggcggtatata cagacccggc 1920
aacaacgtgt tocacacca ggcggcgctg ctgctgggg ccgagcacgt ggcacaccgc 1980
tacaggtgag acatcccat cggcgccggc atctgcgcca gctacacac cgtgagcctg 2040
ctgcggagca ccagccagaa gacatctgt gcttaacca tgaacctggg ccgcagacag 2100
agcatcgctt acagcaacaa caccatcgcc atccccacca acttaagcat cagcatcacc 2160
accgaggtga tgcocgtgag catggccaa gacacgctgg actgcacat gtacatctgc 2220
ggcgacagca ccgagtgccc caactctgt ctgcagtacg gaactttctg caaccagctg 2280
aacggggccc tgaagggcat ccgcgcagag caggaaccga aacccggga ggtgttcgcc 2340
caggtgaagc agatgtacca gacccccacc ctgaagtact toggcggtt caacttcagc 2400
cagatcctgc ccgaccccc ctgaagccacc aagcggagct tcatcgaggc cctgctgttc 2460
aacaaggtga ccttgccgca cgcggcttc atgaagcagt accgagagt cctggcgcac 2520
atcacgcgcc gggacctgat ctgcgcacag aagttcaacg gctgacagt gctgcacccc 2580
ctgctgaccg accagatgat ccgcgcctac accgcgcgcc tggtagcgg ccacgcaccc 2640
cccggttga ccttcggcgc cggcgccgac ctgcagatcc ccttcgcctt gaagatggcc 2700
taaccgttca accgcacgag ctgacccag aactgtctgt acgagacca gaagcagatc 2760
gcacacaggt toacaaagc catcagccag atcaaggaga gctgaacac caccagaccc 2820
gcccgtggca agctgcagga cgtggtgaac cagaaagccc aggcctgaa caactgtgt 2880
aagcagctga gcaacaaact cgggcacatc agcagctgca tgaacacat cctgacccg 2940
ctggacaaag tgaagccga ggtgcagatc gacccgctga tcaaccggcc gctgcagaga 3000
ctgcagacct acgtgacca gacatgata cgggcgcgac agatccgggc ccgcgcacac 3060
ctggccgcca ccaagatgag cgcgtgctg ctgggcaga gcaagcgggt ggaactctgc 3120
ggcaagggct accaactgat gagcttcccc caggccgcgc cccacggcgt ggtgttctct 3180
caagtgaact acgtgcccag ccaggcgagg aacttaacca ccgcgccccc catctgcacc 3240
gagggaaagg cactactccc ccggggaggg gtgttcgtgt taaccggcac cagctggttc 3300
atcacccagc ggaacttctt cagcccccag atcatacca ccgacacac cttcgtgagc 3360
ggcaactcgc acgtgtgtat cggcatcacc acaaacaccy tgcacgccc cctgcagccc 3420
gagctggaca gcttaagga ggaagtggac aagtaactca agaaccacac cagcccccgc 3480
gtggacctgg cgcacatcag ccgcatcacc gccacgctgg tgaactcca gaaggagatc 3540
gaccgcgtga accgagtgcc caagaaactg aacgagagcc tgatgcacct gaaggagctg 3600
ggcaagtacg agcagtcact caagtggccc tgg

```

<210> SEQ ID NO 32

<211> LENGTH: 2094

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized soluble TPA-S1 protein

<400> SEQUENCE: 32

atggacgcga tgaagcagag actgtgcgc gttttgtgc tgtgcggcgc agttttgtc 60
 agtccatcgc cccgggggtc gggatctgac ctatagatagat gcacgacatt cgaatgcgtg 120
 caggcaacca attacaccca acatacttca tcaatgcgcg gcgtttacta tcccgacgaa 180
 atcttccgga gtgacacccat gtatctgact caggacattt ttctgccatt ctacagcaat 240
 gtgacaggct ttacacccat taaccatacc ttogggaacc cagtaaccoc tttaaggat 300
 gggatttact ttgtctgac tgagaaaagt aatgttgtca ggggtgggt ttttggtca 360
 acaatgaaca ataagtctca gagtgtcacc atcatttaaca attctacaca tgtagtacc 420
 agagcatgca acttcagctc ctgtgataac ctttcttttg ctgtgtctaa gcccatgggc 480
 actcaaacac ataactgat ctctgacaaat gcgttcaatt gtacatttga gtatatata 540
 gcgcctctca gccatagcgt ctgggaaaag tccggaaact ttaaacacct gcgggaattc 600
 gtgtttaaga acaaaagatgg attttttgac gtatacaagg gttatcacgc tatcgatgtc 660
 gtgcgtgacc tgcctccggc ctccaacacc ctgaagccta tattcaaat acccctaggg 720
 atcaaacata caaattttag ggcataact acggcatttt ccccgaccca ggaactctgg 780
 ggaacttcgc ccgtgcctca cttttggggc tatctcaagc ctactacatt catguttaag 840
 tatgatgaga atggcacaaat caagatgaca gtggttgct ccagaaatcc actgtctgag 900
 ctgaatgtct ccytaaagag ctctgaatt gataaaggaa ttatatagac cagcaacttc 960
 cgggtcgtgc cctctggcga cgtgtccgg ttcccaaca tcccaaacct ctgcacattc 1020
 ggcagaggtgt tcaacgtcac aaaaatccca agtctctacc cctgggagag gaaaaagctc 1080
 tctaattgtg tggcagatta ttccgtgtta tacaacagca cttctcttc aacgtccaag 1140
 tgttatggcg tgagcgcacc caagottaac gacctctgct tctcaaatgt ataogtgac 1200
 tcttttgtg ttaaggggga cgaatgtcga cagatgcgcc cggggcaaac cggagtgatt 1260
 gcggactaca actataaact gccgcacgat ttactgggtt gtgtgtgttg ctggaatacg 1320
 aggaacattg acgcaacgag ccacgggaac tataattaca aatatgtta cctgcgccat 1380
 ggggaacctca gaccttttga acgagatatt agcaacgtcc cttctcacc ggaatggaa 1440
 cctgtgaccc cactgcacct gaaactgtat tggcctctca acgactacgg cttctaacat 1500
 accacaggga tcgggtacaa gccactatgc gtgtgtgttc tctccttga actcctaat 1560
 gctccgcgca ctgtgtgtgg gccgaagttg agtactgact taataaaaa tcaatgctga 1620
 aactttaact ttaatgctt gacaggtaca ggtgtgtcca caacagatg caaaggttca 1680
 cagccatttc agcaatttgg cagagatgtg totgaattta cagacagcgt gcgcgactcc 1740
 aagactcttg agattttaga catctaacct tgttcttttg gaggagtga cgtgacaact 1800
 ccgggtacca acgcctcacc cgaagtgcct gtctgtgacc aggaagttaa ttgcaacgat 1860
 gtctctacag ccaattcaagc agatcagctg acaacagatt ggcgcactca cagtaacgggt 1920
 acaactgttt tccagactca gcgcggttgt ctgattggcg ccagacacgt ccagacactc 1980
 taagagtgcg atattcccat aggtgcgcgg atttgtgcga gctacacacac tgtataactg 2040
 ctgagagaac caagccagaa atcaattgtg gcatcacaca tgtccttggg agca 2094

-continued

```

<210> SEQ ID NO: 33
<211> LENGTH: 2091
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Uniform optimization of soluble TPA-S1 protein
<400> SEQUENCE: 33

atggagccca tgaaggggg cctgtgtgc gtgtgtgtg tgtcggggc cgtgttgtg 60
agccccagcg cccggggcag cgcagagac ctggaocggt gcaacaactt cgaagcagt 120
caggccccca actacaccca gcaacacagc agcatgggg ggtgtacta cccgagag 180
atctctccga ggcacacact gtaactgacc caggacctgt tctgacact ctacagcaac 240
gtgacccggt tccacacat caacacacac ttgcgcaac ccgtgatccc cttaagga 300
ggcatctact tcgcgcgcc cagaagagc aactggtgc ggggtgggt gttcgcagc 360
aocatgaaca acaagagcca gagcgtgat atcatcaaca acagaccaa cgtggtgat 420
cgggctgca actctagct gtgcagac cccttcttc ccgtgagca gccatgggc 480
acccagaccc acacactgat cttagacaac gccttcaact gcaacttca gtacatagc 540
gacgccttca gcttgagct gagcagagc agcgcaact tcaagcaact ggggagttc 600
gtgtcaaga acaaggagcg ctctctgtac gtgtcaagg cttacagcc catcgactg 660
gtcggggacc tgcccagcgg cttaacacac ctgaagcca tottaagct gccctgggc 720
atcaacatca ccaacttcg ggcatactg acgccttca gccgcgcca ggacatctg 780
ggcaccagcg ccgcgccta ctctgtggc tacctgaag ccaacactt catctgaag 840
tacgcagaga acgcaacct caccagcgc gtgactgca gcagacccc cctgcaccg 900
ctgaagtga cgttgaagc ctctagatc gcaaggcca totaagac cagcaactc 960
cgggtgtgtc caagcgcgca cgtgtgcg ttcccaaca tcaacaact gtgcacact 1020
ggcaggtgt tcaagccac caagtcccc agcgtgtac cttggagcg gaagaagtc 1080
agcaactgag tggcgcacta cagctgtgt tacaagaca cttcttcag caacttaag 1140
tgtacggcg tgaagccac caagctgaac gactgtgtg taagcaact gtaagcaac 1200
agctctgtg tgaaggcgca cgaactgagc agactggcc ccggcagac cggcgtgatc 1260
ccgactacta actacaagt gccgcagac ttcatgggt cgtgtgtgc ctggaacac 1320
cggcaactcg acgcacacg caccggcac tacaactaca agtaacgta cctgcggac 1380
ggcaagctgc gccctctga cgggacatc agcaactgc cttacagcc cgaagcagc 1440
ccctgaccc ccccccact gaactgtac ttgcacctga acgactacg ctctacacc 1500
accaaccgca tcgcgtacca gccctacgg gtgtgtgtg tgaacttga gctgtgaac 1560
gcccccacca cgtgtgtgc ccccaagct agcacagac tgatcaaga cagtgctgt 1620
aacttcaact tcaacggct gaccggcac ggcgtgtga ccccagcag caagcgttc 1680
cagcccttc agactctgc cgggacgtg agcaactca cgaagagct ggggacccc 1740
aagacacagc agatcttga catcagccc tgaacttc gggcgtgag cgtgatcac 1800
ccgggcacca acgcagcag cgaagtgcc gtctgtacc aggactgaa ctgcaccag 1860
gtgagcaccg ccatcacgc cgaacagct acccccact ggcgatcta cagcaacgg 1920
aacaactgt tcaagacca ggcgggtgc ctgactggc ccgagcaact ggacacagc 1980
tacgagtgc acatcccat cggcgccgc atctggcca gtaaacacac cgtgagcgt 2040

```

-continued

ctgcggagca ccagccagaa gagcatgtg gcaacacaa tgagcctggg c 2091

<210> SEQ ID NO 34
 <211> LENGTH: 1623
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Fully optimized soluble TPA-S2

<400> SEQUENCE: 34

atggatgaa tgaagaagg cctgttgtt gttctgtgc tgttggggg ggtatttgg 60
 agtcctctg ccaggggaa cggcagacg agtatagcct actcaacaa taccatgcc 120
 attcctacaa attttccat ctcaatcac acggaagta tgcagttag catggccaa 180
 acctctgtc actgcaacat gtacatctc ggagactcta ctgagtgcg aaacctgtc 240
 ttgcagtat gctcgtttg caccagttg aatcggggc tcagtggcat tgcgcagaa 300
 caagatcga ataccaggga ggtcttcgc caagtcaag agatgtaca aacctctca 360
 ctcaaatact tgggggggt caactttag caaatctgc cagacccct caagctact 420
 aagcgcagt ttatcgaa cttactctt aataagtgga cattagctga tgcggatc 480
 atgaagcag acgagagtg cctgggggat atcaacgcg gggacctaat ctgtgccac 540
 aagttcaag gtctgacag gttccgctt ctctgaccg atgatattg cgcagttac 600
 acgcgcgac tggttagtgg tacggccaa gcaggctga ccttgggtg cgggtctgc 660
 ctgcaaatc catctcgat gcagatgga tacagattt acggcattg agtcccaag 720
 aatgtctat acgagaacca gaagcaaat ctaaacagc tcaacaaga catatccag 780
 attcaggag ccttactac aaccagtac gtttagtga aactgaaga tgtagtgaac 840
 cagaacgtc aggccttaaa tacccttgt aaacagctat cctcaactt tggggctac 900
 tctcctgct caacgatat cctgagcgc ctcgataag tggaaaggga ggtccagatc 960
 gatagactt ttacaggcag gttcagtct ctccagact atgtcacaa acagatcatt 1020
 cgtgtgag agatccgcg ttccgcacaa ttgctgcaa caaagatgtc tgaattgtg 1080
 ctgggaaga gcaagagag gacattttg ggaagagct atcaattgat gagctcccc 1140
 caggccgcc ccaatgaggt ggtattcta acctgagct acgttcacc tcagaacga 1200
 aattcacca ccgcaactgc catttgccac gaagggaag cttatttccc tcagagggc 1260
 gtgtcgttt ttaacgggac tctatggtt ataactcaa ggaattctt ctgcgccag 1320
 ataattcaa cagacaacac ttittgagc ggcatttgc acgtggtcat aggtattatt 1380
 aataatctg tctatgacc gctcagccc gaactggaca gctttaaga ggaagtggac 1440
 aaatactta agaatcata ttaccgcgac gtgactctg gcgacatct cggatcaat 1500
 gctctgttg taaacattca gaaggagac gatcgttga acgaagtggc taagaatctg 1560
 aatgaatcat tgattgacct tcaggagttg ggcagtatg agcagatat taaatggcca 1620
 tgg 1623

<210> SEQ ID NO 35
 <211> LENGTH: 1623
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Uniform optimization of TPA-S2 protein

-continued

<400> SEQUENCE: 35

atggagccca tgaagcgggg cctgtgctgc gtgctgctgc tgtgoggcgc cgtgttcgtg 60
 agcccccagc cccggggcag cggcgacagc agcatgctgc acagaaacaa caacatcgcc 120
 atccccacca acttcagcat cagcatcaac accagagtgga tgcctgtgag catggccaaag 180
 accagcgttg actgcaacat gtacatctgc ggcgacagca ccgagtgccg caacctgtctg 240
 ctgcagtaag cgaagctctg caccacagctg aacccggccc tgagoggcat cgcgcgcgag 300
 caggaccgga acaaccggga ggtgttcgac cagtgaaac agatgtacaa gaccccccac 360
 ctgaagtaat tccgcgcttc caacttcagc cagatcctgc ccgaccccat gaagcccacac 420
 aagcggagct tcactgagga cctgtgtgtc aacaaagtgga cctgggcgca cgcgcgcttc 480
 atgaagcagt accgagagtg cctgggcgac atcaaacgcc gggacatgat ctgagcccaag 540
 aagttcaacg gctgcacagt cctgcccccc ctgtgaccc acgacatgat cgcgcgctac 600
 accgcgcgcc tgggtgagcg caccgcccac gccgagctga ccttcggcgc cggcgccgcc 660
 ctgcagatcc ccttcgcatc gcagatggcc tacggtgtca acggcatcgg cgtgaacccag 720
 aacgtgtctg acyagaacca gaagcagatc gccaacccgt tcacacaggg catcagccag 780
 atccaggaga gctgcacacc caccagcaac gccctgggca agctgcagga cgtggtgaac 840
 cagaacgcc ccggcctgaa caccctggtg aagcagctga gcaacaaatt cggcgccatc 900
 agcagcgtga tgaacgacat cctgaagcgg ctgacaaggt tggagggcga ggtgcagatc 960
 gaccggctga tcaccggccg cctgcagaga ctgcacagat acgtgaacca gaagctgac 1020
 cgggcgcgag agatccgggc cagcgccaac ctggccgcca caaagatgag caggtgctg 1080
 ctgggcgaga gcaagcgggt ggaacttcgc ggcgaaggct accaactgat gagcttcacc 1140
 agggccgcgc ccaacggcgt ggtgttcctg acgtgacat acgtgccag ccaggagcgg 1200
 aacttcacca ccgcgccccc catctgccac gagggcaagg cctacttccc ccgggagggc 1260
 gtttcttgtgt tcaacggcac cagctggttc atcacccaga ggaactctt cagcccccag 1320
 atcatcacca ccgaacaacac ctctgtgagc ggcactgctg acgtgtgat cggctatcat 1380
 aacaacaccc tgtacgaccc cctgcagccc gagctggaca gottcaagga ggaagctgac 1440
 aagtaactca agaaacacac cagcccccag ctgaacctgg cgcacatcag cggcatcaac 1500
 gccagcgttg tgaactacca gaagagatc gaccggctga ccgagctggc caagaacctg 1560
 aacgagagcc tgatcgacct cgaagagctg ggcagatag agcagtaaat caagtgccc 1620
 tgg 1623

<210> SEQ ID NO 36

<211> LENGTH: 1269

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized N protein

<400> SEQUENCE: 36

atgtccgata atgttcccca gtotaaaccag aggtcggcgc caagaatcac attccggggc 60
 ccaacagaca gtaccgataa caaccagaac ggcggaagaa accggggccag gcccaagcag 120
 cggagacctc agggattacc aaataatacc gcaagctggt tcacagccct gaccagcat 180
 ggaagaagag aactgagatt ccttagagga caaggggtgc ctattaatat taatagcggg 240

-continued

```

cctgacgac aaattggcta ttatcgacgt cgcactcgcc gtgttagagg gggggacggg 300
aagatgaag agcttagccc acgctggtac ttttactatc tgggaacccg acctgaagct 360
agctcgccct acgctgctaa caaggaggga atagtatggy tgcgcacgga aggtgcgttg 420
aatactcga aagatcacat cggcaccaga actcctaaca ataacgcgcg aaccgctcta 480
caattacccc agggaaactac tctgcgaag ggyttctatg cggagggaag ccgcgcgcgc 540
tcacaagcca gttaacgtct cagctcccg tgcaggggta attcccgaaa cagcaacccc 600
ggatcatcta ggggaactc tccgcgccg atggctcag cggcgggcga aacagctctg 660
gctctgctat tgcctgacgc gctcaaccag ctgagtcaca aagtctcttg taaggtcag 720
cagcagcagg gtaaacagt gaccaaaaa agtcagcgcc aggcacgcaa gaacacacgc 780
cagaacagta cggccacaaa gcaatacaat gtgacccaag cttttggaag cggggggccc 840
gaacagacac agggcaattt cggcgatcaa gatttgatca gacagggcac tgactacaaa 900
cactggcgcc agatcgctca gtttgcacct agcgcctccg ctttcttgg catgagtgg 960
attggcatgg aggtgacacc atcaggtact tgyttaacgt accacggggc aatcaaaatt 1020
gatgataaag atcccagtt taaggacaac gttatcctcc tgaataagca tattgacgcc 1080
tataagacct tcccccaac cgaacaaaag aaggaacaga agaagaagac agacagagca 1140
cagcctctcc cccagagga gaaaagcag cctactgtca ccttctgcc cgtcgcagac 1200
attgatgact ttcccgcca actccagaac tctatgagtg ggcctccgc tgactctacg 1260
caggcctga 1269

```

```

<210> SEQ ID NO 37
<211> LENGTH: 1266
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Uniform optimization of N protein
<400> SEQUENCE: 37

```

```

atgagcgaca acggccccc gagcaaccag agaagcgccc cagaatcac cttcggcgcc 60
cccacgcgca gcaccgacaa caaccagaac ggcggcagaa acggcgccag acccaagcag 120
agaaagacccc agggcctgccc caacaacacc gccagctggt tcacgcacct gacccagcac 180
ggcaagyaag agctgagatt ccccagagcg caggggctgc ccatcaacac caacagcgcc 240
cccagcgacc agatcgctca ctacaagaag gccaccagaa gagtgaagag cggcgacggc 300
aagatgaag agctgagccc agatgggtac ttctactacc tgggcacccg ccccagggcc 360
agcctgccct acgctgcaca caaggaggcg atcgtgtggy tggccacaga gggcgccctg 420
aacaacccca aggaacacat cggcaccaga aacccaaca acaacgcgcg caccgtgctg 480
cagctgcccc agggcaccac cctgcccaag gcttctacg ccgagggcag cagagcgccg 540
agccagacca cagcagaag cagcagcaga agcagaggca acagcagaan cagcaacccc 600
ggcagcagca gaggcaacag ccccgccaga atgcgcagcy gggcgggcga gacgcgccg 660
gcccgtctgc tgcctgacag actgaaccag ctgagagcga agytgagcgg caaggccag 720
cagcagcagg gccagacgtg gaccagaag agcgcgcgcg aggcacgcan gaagcccaga 780
cagaagagaa ccgcacacaa gcagtacaac gtgaccacag ccttggcag aagagcggcc 840
gagcagaccc agggcaactt cggcgaccag gacctgatca gacagggcac cgactacaag 900

```

-continued

```

oactggcccc agatgcccc gttagcccc agcgccagcg ccttcttcgg catgagcaga 960
atcgcatcgg aggtgacccc cagcgccacc tggctgacct accacggcgc catcaagctg 1020
gacgacaaag acccccagtt caaggacaaac gtgactctgc tgaacaagca catcgagcgc 1080
tacaagacct tccccccacc cgaagccaaag aaggacaaga agaagaagac cgaagaggcc 1140
cagccccctgc cccagagaca gaagaagcag cccacagtga ccttcttcgc cgcgcgcgac 1200
atggacgact tcagcagaca gctgcagaac agcatgagcg gcgcccagcgc cgaacagacc 1260
caggcc 1266

```

```

<210> SEQ ID NO 38
<211> LENGTH: 1209
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Fully optimized N protein lacking NLS

```

```

<400> SEQUENCE: 38
atgagtata atggccccca gtctaaccaag aggaagcgac cggggtacac gttcgttggc 60
ccacccgact caacagacaa taatcagaac ggaaggcgca atggtgcacg tcttaagcag 120
agaagccccc aagggtctgc taataataca gcaagtgtgt ttacgcact cacaacaacat 180
ggaagaagag agttgcggtt cccccggcgc cagggtctgc ccatcaaac aaatagcgga 240
ccgacgcac agatcgcata ttaccgaaga gctacaagga gagttcggc cggggatggc 300
aagatgaag agctatcacc acgatgttac ttctattacc tcgggaagag cccagagcgc 360
tcgtaccat acggggccaa caaggagggt attgtctggg tcgtatccga agggccctg 420
aatacacota aagaccacat aggtaccaga aatcccaaca ataacgcgc gacggtgta 480
cagctctctc agggaagcac cttccaaaaa gggttttacg ccgaaggatc tcggggaggg 540
tcacaggata gctcccgtag ctctccaaag tccaggggga attctagaaa cagtacacc 600
ggctctagcc gtgttaactc cccagctcgc atggcatccg ggggagggga aaccgctctg 660
gctctgtctc tgttatagtg gttgaacaaa ctgaatcga aggtatccgg aaaggagacg 720
cagcagcaag cccagactgt gactaagaag tccgcggcgc aggcacgtaa gaacccccgc 780
cagaacagaa ctgcaccaaa acagtataat gtgacacagc ccttcggcag acggggtcca 840
gagcagaccc aaggcaactt cggggatcag gacctgattc ggcagggtac cgaactaag 900
cactggccgc aaattgctca gtttctccc agtgcagtg ccttcttcgc catgtctagg 960
atcgggatgg aggttactcc tagcggcact tggcttact atcacggagc catcaaacct 1020
gatgataagg accacagtt taaggataac gtgattctgc tgaacaacaa tatagacgcg 1080
taccctctcc cgcgaaggca gaaaaaacag cttacgctca cgttactgcc tgcgcagac 1140
atggacgact ttctagaca gttgcacaaa agcatgtcag ggcgcatccg cgaatgacat 1200
caagottga 1209

```

```

<210> SEQ ID NO 39
<211> LENGTH: 1206
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Uniform optimization of N protein lacking NLS

```

```

<400> SEQUENCE: 39

```

-continued

atgagcgaca acgcccacca gagcaaccag agaagcgccc cagaatcac cttggcggc 60
 cccaccgaca gcaccagaca caaccagaa ggcggcagaa acggcgccag acccaagcag 120
 agaagacccc agggcctgccc caacaacac gccagctggt taaccgccc gaccacgac 180
 ggcagaggag agctgagatt cccagaggc caggggctgc ccatcaacac caacagcggc 240
 cccgacgacc agatcggtca ctacagaaga gccaccagaa gactgagagg cggcgacggc 300
 aagatgaagg agctgagccc agatgggtac ttctactacc tgggcaccgg ccccgaggcc 360
 agcctgccc acgcgccaaa caaggagggc atcgtgtggg tggccaccga gggcgccctg 420
 aacaccccaca aggaaccacat cggcaccaga aaccccacaa caacgcgcgc caccgtgctg 480
 cagctgcccc agggcaccac cctgcccaag gcttctaac cgaaggcgag cagagcggc 540
 agccaggcca gcagcagaag cagcagcaga agcagggcca acagcagaaa cagacccccc 600
 ggcagcagca gaggcaacag cccgcagaa atggccagcg gcggcgcgga gaccgccc 660
 gccctgctgc tgcctgacag actgaaccag ctggagagca aggtgagcgg caaggcgag 720
 cagcagcagg gccagaccgt gaccagaag agcgcgccgg aggcagcaga gaagcccaga 780
 cagaagagaa ccgcaccaca gcagtacaac gtgacccagg ccttcggcag aagagccccc 840
 gagcagcccc agggcaactt cggcgaccag gacctgatca gacagggccc cgaatcaag 900
 cactggcccc agatcgccca gttccgcccc agcgccagcg ccttcttcgg catgagcaga 960
 atcgccatgg aggtgacccc cagcggcacc tggctgacct accagcgcg catcaagctg 1020
 gacgacaagg acccccagtt caaggacac gtgactctgc tgaacacgca catcgacgcc 1080
 taacccctgc cccagagaca gaagaagcag cccaccgtga cctgtgccc cgcgcgcgac 1140
 atggacgact tcagcagaca gctgcagaa agcatgagcg gcgcagcgcg cgacacgacc 1200
 caggcc 1206

<210> SEQ ID NO 40

<211> LENGTH: 666

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> OTHER INFORMATION: Fully optimized M protein

<400> SEQUENCE: 40

atggctgaca acggcacccat aaccgtcag gagcttaaac agttattaga acaatggaa 60
 ttgggtatag gattccctatt tctggcatgg atcatgctgc ttcaagttgc ctattctaac 120
 cgaataggt tttgtacat tatcaagctg gtcttccctt ggctgctcg gcccttaaca 180
 ctacgcctgtt ttgttttggc ggcctgtgat cggatcaatt gggtagacag tggcattgct 240
 attgcgatgg cttgcactgt ggggctgatg tggctgtcgt atttcgtgc ctactcccg 300
 ctgtttgccc gaacaaggag tatgtggtct tttaaccccg agaccaatat tctgtcaat 360
 gtgcctttac gggcactat cgtgaccgg cctctaattgg aatccagct ggttaattggc 420
 cgaatcatca taagggggca cctcagaatg gccgggcacc caactgggag atgcgacatc 480
 aaggatctgc cgaaggaaat taactgttca acttaacgaa cgtgagcta ttacaactg 540
 ggagctagcc agagagtggg taaccgactcc ggccttcgtc cctacaacgc ctacgtatc 600
 ggaaattaca aactcaacac agatcatgca ggaagcaatg ataactcgc cctcctgtgc 660
 cagtga 666

-continued

<210> SEQ ID NO 41
 <211> LENGTH: 663
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Uniform optimization of M protein
 <400> SEQUENCE: 41
 atggccgaca acggcaccat caccgtggag gagctgaagc agctgctgga gcaetgggac 60
 atggctgacg gcttctctgt cctggcctgg atcatgctgc tgcagtttgc ctacagcaac 120
 agaaacagat tctgtacat catcaagctg gtgttctctg ggctgctg atcgctgacc 180
 ctggcctgct tctgtctggt cgcctgtgac agaataaact gggtagccgg cggcctgccc 240
 atcgccatgg cctgcactgt gggcctgatg tggctgagct acttctgtgc cagcttcaga 300
 ctgttcgcca gaaccagaag catgtggagc ttcaaccccg agaccaacat cctctggaac 360
 gtgcccctga gaggaccat cgtgaccaga cccctgatgg agagcagct ggtgatctgc 420
 gccgtgatca tcagaggcca cctgagaatg gccggccacc cctggggcag atcgacatc 480
 aaggacctgc ccaaggagat caccctggcc accagcagaa ccttgagcta ctacagctg 540
 ggcgcagccc agagctgggg caccgcagac gcttctgccc cctacaacag atacagaatc 600
 ggcactatac agctgaacac cgaccacgcc ggcagcaacy acaacatcgc cctgctgggt 660
 cag 663

<210> SEQ ID NO 42
 <211> LENGTH: 231
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Fully optimized E protein
 <400> SEQUENCE: 42
 atgtacagct ttgtgtctga agaaacagga acygtgatag ttaatagtgt ttgtctttc 60
 ttacgcttgc tagtcttctc tctgtcaca ctggccattt taactgctgt tctgtatgc 120
 gcttactgtt gcaatctcgt aaacgtgtgc ctgtttaaac caacggttta cgtatactgc 180
 cgagttaaaa accctgaattc ttacagaagg gtctctgata tgcagtcta a 231

<210> SEQ ID NO 43
 <211> LENGTH: 231
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Uniform optimization of E protein
 <400> SEQUENCE: 43
 atgtacagct tctgtgagca ggaagccggc accctgatcg tgaacagcgt gctgctgttc 60
 ctggccttgc tgggttctct gctggtgaac ctggccatcc tgaacgccct gggctgtgct 120
 gctactctgt gcaacatcgt gaacgtgagc ctggtgaagc caacgtgta cgtgtacagc 180
 cgggtgaaga accgtgaacag cagcgagggc gtgcacgccc tctgtgtgtg a 231

<210> SEQ ID NO 44
 <211> LENGTH: 3588
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
 <223> OTHER INFORMATION: Minimal optimization of soluble S protein
 <400> SEQUENCE: 44

atgtttatct tctgtgtgtt tctgacactg acaaagcgga gtgacctgga tagatgcaca	60
acgtttgacg acgtgcagcg ccccaactac accacgcata catccagcat gegggcggtt	120
tactaccccg atgagctctt tagaagtgtat actctgtatc tgactcagga cctgtttctg	180
cocttotatt ctaaagttac tggcttccat acaatcaacc acaccttogg caaacccgta	240
atacacttta aggtaggcat ctactttggc gccacggaga agtctaaagt agtgagggc	300
tggtgtgttg gcagtactat gaacaaacag tctcagctct tgataaata caacaaactc	360
actaacgtcg tcatacagag cgtgaacttc gagctgtgcy ataaccctt cttggccgtt	420
tcgaagccca tgggcaactca gccactatac atgacttttg ataagcctt caactgcacc	480
tttgagtata tctctgatgc cttcagctcg gatgtgtccg agaagtcagg caactcaag	540
catctgagag agtttgtgtt caagaaacag gatggcttcc tgtaactata caagggctac	600
cagcccatag atgtgtgtac tgacgtgcc agcggttcca acactctgaa gccactattc	660
aagctgcccc tgggcataaa cattacaacc tttagagcca ttctgacggc cttctcccc	720
gccacagata tctggggcac aagtgccgcc gccactcttg tgggtacct gaagcccaac	780
acttttatgc tgaagtacga cgaagaacgc accataacag atgcccgtga ctgttctcag	840
aaccocctgg ccgagctgaa gtgcactgtt aagagttttg agatagataa gggcatctat	900
cagacaagca acttcgcgtt ggtcccagc ggcgatgttg tgaggttccc caacattacc	960
aaactgtgcc ccttggcgga ggtattcaac gccacaaagt tccctccgt ttacgctbgy	1020
gagaggaaga agatttcaaa ctgctgtggc gactactcgy tctgtataa ctctacttc	1080
ttcagtaact ttaagtcta cggcgtgtct gccacaaagc tgaacgatct gtgctttagc	1140
aacgtgtatg ccgatagcct cgtctcgaag ggcagcagc tcagacagat cgcocccgga	1200
cagacaggcg tcactgcga ctacaactac aagctgcocg acgatttcat gggctgtgctg	1260
ctggcctgga acacaggaaa catagatgco accagcactg gaaactacaa ctacagtao	1320
agatatctgc ggcacggcaa gctgagggcc ttgagagag acatctctaa cgttccctt	1380
tcocccagtg gcaagccctg cactccccc gccctgaact gctactggcc cctgaacgac	1440
tatggcttct acaccacaa tggcatoggc tctacgcctt accgcgtagt cgtgtgtcgy	1500
ttcagagctgc tgaagccccc cgcacagctc tgcggcccca agctgtccac tgactgtatt	1560
aagaacagat gttggaactt caactttaac ggcctgaact gaacccggct gctgacacc	1620
agcagaacga ggttccagcc cttccagcag tttygcagag acgtgtctga ttccacagat	1680
tcctgtgagag atcccaagac ttccagata ctgatatca gtccctgctc cttcggcgga	1740
gtgtcagtta ttacaccggc caactaaagc tcgtccaggy tagccgttct gtatcaggac	1800
gtgaactgca ctgatgtgag tacagcaatc caagccagac agctgaaccc cgcctggggy	1860
atttatgta cgggcacaaa cgtctttcag actcaggccg gctgcctgat cggcgcggag	1920
catgtagata cgtcttatga gtgcacatc cccactggcg ccgcatctcg cgcacagat	1980
caacacgttt ctctgctgcy aagtaactct cagaagtcta tagtggcta caacatgtct	2040
ctggggcccg atagctctat cgcctataga acaacaacta tagccatccc caacaacttc	2100
tcattttcta tcactacaga ggtgatgccc gtctccatgy caagacagc cgttgatgbc	2160

-continued

```

aacatgtaca tctgcggcga tagtacagag tgcgcacaac tctgtctgca gtatggcagc 2220
ttotgcaccc agctgaacag agccctgtct ggcctgcgcg ccgagcagga taggaacaca 2280
agagaggttt tcgccagagt taagcagatg tacaaagactc ccaacttgaa gtactttggc 2340
ggctttaact ttctcagat tctgcocgat cccctgaagc ccaactaagag gagtttcata 2400
gaggacctgc tgttaacaaa ggtgactctg gccgaagccg gctttatgaa gcaatagcgc 2460
gagtgcctgg gcgatatcaa cgcagagagc ctgatctgtg ccagaagtt taacggcctg 2520
acagtactgc ccccccctgt gactgatgac atgattgcgc cctatacggc cgccctgggtg 2580
tctggcaact ccacgcgcg ctggaccttt ggcgcgcgcy cgcccttgca gatacccttt 2640
gccatgcaga tggcctaccg attcaacggc ataggcgtaa ccagaacgt tctgtatgag 2700
aaccagaagc agatagccaa ccagttcaac aaggccatct ctcagattca ggagtctctg 2760
aocactacat ctactgcctt gggcaagctg caggacgtag tgaaccagaa cgcocaggcc 2820
ctgaacaccc tggttaaaga gctgtcaagt aacttggcgc ccaactctga cgttctgaac 2880
gatatactga gtccgctgga taagttggag gccagagtgcc agattgacag actgatacaa 2940
ggcagactgc agtctctgca gacatatgtt actcagcagc tgataagggc cgcgcagatt 3000
agagccagtg caaacctggc cgcacataag atgtccagct gcgtcctggg ccagagtaag 3060
agggtagact ttgttgcaa gggctatcac ctgatgtcct tccccaggc cgcoccccac 3120
ggcgtcgtgt ttctgcattg cacttatgtt cctcaccagg agaggaactc cacgaacgcc 3180
ccgcacatct gccacgaggy caaggcctat ttccccaggy agggcgctct cgtattcaac 3240
ggcacagatt ggttcattac ccagcgaaac ttcttttcgc ccagataat tacaaaggac 3300
aacaactttg taagtggcaa tctgcagctg gtcattggca taataacaaa cacggtttac 3360
gacccctgcg agcccgagct ggaattcatt aagagagagc tggacaagta cttcaagaac 3420
catactagcc ccgacgttga tctgggcgac ataagoggca tcaacgccag tctagtcaac 3480
atacagaagg agatcgatag actgaacgag gtggccaaga acctgaacga gtctctgata 3540
gaactgcagg agctgggcaa gtacgagcag tacatcaagt ggcctctg 3588

```

```

<210> SEQ ID NO 45
<211> LENGTH: 2049
<212> TYPE: DNA
<213> ORIGIN: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Minimal optimization of soluble S1 protein
<400> SEQUENCE: 45

```

```

atgttcactt tctgtctgtt tctgacactg actctgtggt cagatctgga tagatgact 60
aactttgacg atgtacaggc ccccaactac actcagcaca catcytccat gcgagcgtg 120
tattaccocg acgagatctt cagaaagtgc actctgtcac tgacaagga cctgttctctg 180
cccttttact ctacagtgac tggctttcac actatacaac atactctcg caaccccgta 240
atcccttca aggatggcat ctattttgac gccaccgaga agtccacagt ggtgagggc 300
tgggtctctg gcagtcatgt gaacaacaa gtccagtcgc tgataatcat aaacacagt 360
actaacgtgg ttataagagc ctgcacctic gaetgtgcg acaaccctt cttgcgcgtg 420
tccaaagcca tgggcacaca gaaccacac atgatattcg acaacgctt taactgtact 480
tcgagatata taagcgatgc cttcagctcg gatgtttctg agaagtcagg caactttaag 540

```

-continued

catctgagag agtttgtatt caagaacaag gaacggttcc tgtatgttta taagggtac 600
 cagcccatag atgtgtgtcg ggtatgtccc agcgggttca acacaactga gccattttt 660
 aagctgcccc tgggcatcaa cataccaac tttagagcca tctgactgc ctttagcccc 720
 gccagagata tatggggcac tagcgccgcc gccattttcg tggctacct gaagccacac 780
 acattcatgc tgaagtacga tgaagacggc acatttacgg atgcctaga ttgcgtcag 840
 aacccccctg ccagactgaa gtgcagtgtg aagcttttcg agatcgacaa gggcatatac 900
 cagaattcta actttcgggt ggtcccccag ggcgaagtgt ttggtttccc caacacacc 960
 aacctgtgcc ccttcggcga ggtgtttaac gccacaaagt tccctccgt atatgctgg 1020
 gagaggaaga agatttcgaa ctgctgtggc gactatagcg tctgtacaa ctctacattc 1080
 tttctacat tcaagtgcga cggcgtcagt gccactaagc tgaagacct gtgcttcagc 1140
 aaagtgtat ccgaactcatt tgaagttaa ggcgatgat tgagacagat tgcgcccgcc 1200
 cagacaggcg tgatgcgga ttataactat aagctgcgcg acgaattcat gggctgcgtt 1260
 ctggcctgga acacaaggaa catcgtgcc actagcactg gcaactacaa ctacaagtac 1320
 aggtatctga gaacaggcaa gctgagggcc ttgcagcgag atatcagtaa cgtacccctc 1380
 agtcccgacg gaagccctg cactccccc gccctgaact gctattggcc cctgaacgac 1440
 tacggctttt ataccactac aggcactcgg taccagccct acagggttgt ggtgctgagc 1500
 ttcgagctgc tgaacgcccc cgcactgttt tgcggcccca agctgtcaac ggaatctgac 1560
 aagaaccagt gcgtaaactt taactttaac ggcctgacag gaacaggcgt cctgactccc 1620
 tctagttaga gtttcacgac ctttcacgag ttgcgcgcg acgtacagca ttttaaggat 1680
 agtgtgagag atcccaagac cagcagatgc ctggaactta gtccctgttc ttctcgccgc 1740
 gtgtctgtca taacgcccg caggaacgcc tctttgaggt tgcggcttct gtaccaggac 1800
 gtcaactgta cagaactctc cacagccata caccgcgata agtgactcc cgcctggaga 1860
 atttactcta ccggcaacaa cgtcttcacg accagggcgg gctgcctgat cggcgccgag 1920
 catgtggata cttctacaga gtgcagacata cccatggcg ccggcatttg cgcctgtac 1980
 catacgtgt ctgtgtgag atctacctct cagaagagta tegtgtcta cactattccc 2040
 ctggggccc 2049

<210> SEQ ID NO 46

<211> LENGTH: 1539

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Minimal optimization of soluble S2 protein

<400> SEQUENCE: 46

gatagcagca tagctacta aaacaacaag atcgccatcc caacaaactt ttccattccc 60
 ataactacgg aggtgatgoc cgtgacgatg gccaaagact cgttagattg caacatgtac 120
 atatgtggcg attctacaga gtgtgccaac ctgctgctgc agtacggctc ttctgcaag 180
 cagctgaaca gggccctgtc tgcactgcc gccagacagg atcgaaacac acgggaaggtt 240
 ttgcgccagg taagacagat gtataagacg ccaactctga agtacttcgg ccggttcac 300
 ttctctcaga tactgcggca cccctgaag cccactaaga ggtctttat cgaagatctg 360
 ctgttcacaa aggttacact ggcgatgcc ggtttatga agcagtatgg ccagtgctgt 420

-continued

```

ggcgacatca acgcccagaga tctgatatgc gccacgaagt tcaacggcct gactgtgtgt 480
ccccccctgc tgactgacga catgatgcgc gccatataccg ccgcccctgt gagtggcaaa 540
gccactgcgc gctggacatt cggcgccgyc ggcgcctgc agatccccct cgcctgcag 600
atggcctaca gatttaacgg cattgcyctc actcagaacg tctgtatga gaaccagaag 660
cagatcgcca accagtttaa caaggccata agccagatcc aggaagtaact gacaaacaga 720
agtaaccgcc tgggcaagct gcaggatgta gtgaacaga acgcccaggc cctgaacact 780
ctggttaagc agctgtctag caactcggc gccatcagta gtgttctgaa cgtattctgt 840
cttaggtcgg acaaggtcga ggcgagggtg cagattgata gcctgattac cggcagactg 900
cagagtctgc agacttatgt aactcagcag ctgatcagag ccgcccagat tcgagcctcc 960
gccaacctgg ccgccacaaa gatgtctgag tgcgtcctgg gccagagtaa gagggttgac 1020
ttctgcggca agggctatca tctgatgtct tttccccagg ccgcccccca cggcgctgtg 1080
ttcctgcacg taacttacct gccacgcag gagagaact ttacacactg ccccgccatc 1140
tgccacgagc gcaaggccata cttccccaga gagggcgtgt ttgtgttcaa cggcacatct 1200
tggttcaatca cccagagtaa cttttcagc cccagatcaa taacaactga caaaccttc 1260
gtttcgggca actcgacact agtctcggc ataataaaca acacactgta cgtacccctg 1320
cagcccagc tggacagctt taaggaggag ctgacaagt acittaaaga ccatacctca 1380
ccgatgtgg acctggcgca cattctcggc ataaagcct ccgtcgtcaa catccagaag 1440
gagatagata gactgaacga ggttgccagc aacctgaacg agtccctgat cgtactgcag 1500
gagctgggca agtacgagca gtatataaag tggcctcgg 1539

```

```

<210> SEQ ID NO 47
<211> LENGTH: 1620
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Minimal optimization of TPA-S protein

```

```

<400> SEQUENCE: 47

```

```

atggatgcca tgaagcgagg cctgtgttc gtactgtctg tgtgcgcgc cgtgtttgtg 60
agccccagcg cggcggcgag tggcgacagc agcatcgctt attcgaacaa caactattgc 120
ataccacaaa acttctctat ctctataact acggaggtga tgcocgtgtc tatggccaa 180
actagtgtag actgcacatc gtacatctgc ggcagactca ctgagtgcgc caactgtgt 240
ctgcagtatg gctctttctg caccacgctg aacagagccc tgagtggcat cgcgcgcgag 300
caggacggga acaacagaga ggtttcgcc cagtgaaagc agatgtacaa gacccccact 360
ctgaagtatt ttggcgctt caactctctc cagatcctgc cgtacccctc gaagccacc 420
aagaggtctt tcatcgagga cctgctgttc aacaaggtca cctcggcga tgcgcgttc 480
atgaagcagt acggcgagtg cctggcgagc attaaagccc gcgacctgat ctgtccccc 540
aagtttaacg gctcagcgt cctgcctccc ctgtgcacag atgatgatgt cgcgcctac 600
actgcgcccc tggctctcgg caacgcacc gccgctgga ctttggcgc cggcgccgc 660
ctgcagatcc ccttcgccat gcagatggcc tatagattta acggcatagc cgtaaactcag 720
aacgtcctgt acgagaacca gaagcagatc gccaacaggt taaacaggc catctccacg 780
attcaggaga gctgacaaac cactagcact gccctgggca agctgcagga cgtggtgaac 840

```

-continued

cagaacgccc aggcocctgaa cacactgggt aagcagctga gtttaactt tggcgccata	900
tctcgggtgc tgaacgacat actgtcaagg ctggacaagg tcyaggcoga gggtcagata	960
gatagactga tcacagcgag actgcaagag ctgcgaactt acgtatcaca gcagctgac	1020
agagccgcgc agatcagagc ctacagccaac ctggccgcga cgaagatgic tgagtgcgtc	1080
ctggggcagt ctaagagagt cgaattctgc ggcgaaggct accaactgat gagttccccc	1140
caggccgcgc cccatggcgt tgtattccgt catgtgacat atgttccctc ccaggagagg	1200
aactttacca cyggcccccgc catctgcacg gagggcaagg cctacttccc cagagagggc	1260
gtgttcgttt ttaacggcac tagctgggtt attacccaga ggaacttctt ctccccccag	1320
attataacaa cagataacac ttctcgtgac ggcgaactgc atgttgtgat aggcataatt	1380
aacacacacg tgaacgatcc cctgcagccc gagctggata gtttaagga ggagctgac	1440
aagtatttta agaaccacac ttccccgat gtacacctgc gogatatac tggcataaac	1500
gcacgtgctg tgaacataca gaaggagac gataggctga acgaggtggc caagaacctg	1560
aacaggtcac tgatgatctc gcaggagctg ggcgaagtac agcatatat taagtggccc	1620

<210> SEQ ID NO 48
 <211> LENGTH: 231
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Minimal optimization of E protein
 <400> SEQUENCE: 48

atgtatagtt ttgtgagtga ggaagcgggc accctgattt tcaactcagt gctcgtgttc	60
ctggcccttg ttgtcttctc gctgttaact ctggccatcc tgaactccct gacactgtgc	120
gcctactgct gcaacatcgt gaactctctc ctggtaaagc ccaactgtta cgtgtattct	180
aggggtgaaga acctgaactc cagcaggggc gttcccgatc tgcgtgtatg a	231

<210> SEQ ID NO 49
 <211> LENGTH: 1620
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Minimal optimization of TPA-S2 protein
 <400> SEQUENCE: 49

atggatgcca tgaagcgaag cctgtgtgc gtactgctgc tgtgcggcgc cgtgtttgtg	60
agccccagcg cccggggcag tggcgacaga agcatcgctt attcyaacaa cactattgac	120
ataccacaaa actctcttat atctataact acggaagtga tgcocgtgic tatggcccaag	180
actagttag actgaacat gtacatctgc ggcgaactca ctgagtgcgc caacctgtgc	240
ctgcagtatg gctcttctc caccacagct aacagagccc tgagtggcat cgcgcgcag	300
caggacccga acacaaagga ggttttcgac caggtaaagc agatgtacaa gaccoccaact	360
ctgaagtatt ttggcgctt caactctctc cagatcctgc ccgatccctc gaagcccaac	420
agaggtctt tcatcgagga cctgctgttc aacaaagtca ccttgccga tgcgcgcttc	480
atgaagcagt acggcgagtg cctggcgac attaacgccc gcgacctgat ctgtgcccag	540
aagttaacg gctgacgct cctgcccccc ctgctgacag atgatattgat cgcgcctaac	600
actgcgcgcc tggctctctg caccgcaccc gccgcttga ctttcgggcg cggcgccgcg	660

-continued

```

ctgcagatcc ccttcgccat gcagatggcc tatagattta acggcatagg cgtaactaag 720
aacgtccctgt acgagaacaa gaagcagatc gccaacaggt ttaacaaggc catctcccaag 780
attcaggaga gcctgcaaac cactagcnet gccctgggca agctgcagga cgtggtgaac 840
cagaacgcgc aggcocctgaa cacactggtt asgcagctga gttctaactt tggcgccata 900
tcctcgtgtc tgaacgacat actgtcaagc ctggacaagc tcgaggcgca gggtcagata 960
gatagactga toacaggcac actgcagagc ctgcagacct acgttacaca gcagctgate 1020
agagccgcgc agatcagagc ctoagccaac ctggcccgca cgaagatgtc tgagtgcgtc 1080
ctgggccaagt ctaagagagt cgattctctg ggcaaggcct accaactgat gagttccccc 1140
caggccgcgc cccatggcgt tgtattctct catgtgacat atgttccctc ccaggagagg 1200
aactttacca cggcccccgc catctgccac gagggcaagc cctacttccc cagagagggc 1260
gtgttcgttt ttaacggcac tagctggttt attacccaga ggaacttctt cccccccag 1320
attatacaca cagataaacac ttctgtgtcc ggcaactcgc atgttgtgat aggcatactt 1380
aacacacacg tgtacgaccc cctgcagccc gagctggata gtttaanga ggcctggac 1440
aagttattta agaaccacac ttcccgcgat gtgaacctgc gcgatatcag tggcataaac 1500
gccagtgctc tgaacataca gaaggagatc gataggctga acgagtggtc caagaacctg 1560
aacgagtcac tgcctgatct gcaggagctg ggcaagtatc agcagtatat taagtggccc 1620

```

<210> SEQ ID NO 50

<211> LENGTH: 2052

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Sequence contain in VR9208

<400> SEQUENCE: 50

```

atggttatct tctctgtgtt cctcacctc acccagcgca gcgactgga taggtgcacc 60
aacttcgacg acgtgcaggc ccccaactac acccagcaca ccagcgcgat gagggcgctg 120
tactaacccg acgagatttt cagagacgac accctgtcac tcacccagga cctgttctctg 180
ccctttatca gcaacgtgac cgttctccac accatcaccc acaacttcgg caaccocgtg 240
atccctttca aggacgcgat ctactctcgc gccacacgaa agagcaatgt ggtgcggggc 300
tgggtgttgc gcagcaccat gaacacaaag agccagagcg tgatcatcat caacacagc 360
accaacgttg tgatctcggc ctgcaatttc gagctgtcgc acaacccttt ctctccgtg 420
tcacaaacta tggggcaccca gaccacacac atgatctctg acaacgcctt caactgcacc 480
ttcagatcac tcagcagcgc cttcagcctg gatgtgagcg agaagagcgg caactcaag 540
cacctgcggg agttcgtgtt caagaacaaag gacggcttcc tgactgtata caagggtcac 600
cagcccatcg acgtggtgag agacctgcgc agcggcttca acaccctgaa gccatcttc 660
aagctgcgcc tggggcatcaa catcaccaac ttccgggcca tctcacccgc ctttagccct 720
gccacagata cctggggcac cagcgcgcgt gctactcttg tgggtacctc gaagcctacc 780
accctcatcg tgaagtacga cgaagaacgc accatcacgc atgcctgga ctgcagccag 840
aacccctcgg ccagagctgaa gtgcagcgtg aagagctctg agatcgacaa gggcatctac 900
cagaccagca acttcagagt ggtgcctagc ggcagctggt taggttccc caatataacc 960
aacctgtgac ccttcggcga ggtgttcaac gccacaagt tccatagcgt gtaocctggt 1020

```

-continued

```

gagcggaaga agatcagcaa ctgctgagcc gattacagcg tgcgtacaa ctccacttc 1080
ttcagcacct tcaagtgotc cggctgagc gccacaaagc tgaacagacct gtgcttcagc 1140
aacgtgtacg ccgactcatt cgtggtgaag ggcagcagcy tgagacagat cgccctcggc 1200
cagacccgcyg tgatgcgcga ctacaaactac agcttcccy acgacttcatt gggctgcgtg 1260
ctggctcgtg acaccagaaa catcgacgccc acctccaccy gcaactacaa ttacaagtac 1320
cgctacctga ggcacggcaa gctgagaccc ttcgagcggg acatctccaa cgtgccttc 1380
agccccagcg gcaagccctg caccocccct gccctgaact gctactgccc cctgaacgac 1440
tacggcttct acaccacac ccgcatcggc tatcagccct acagagtgtt ggtgctgagc 1500
ttcgagctcg tgaagccccc tgcaaccctg tgcggcccca agctgagcac cgaactcate 1560
aagaaccagt gcytgaactt caacttcaac ggcctcaccy gcaagcgctt gctcaccccc 1620
agcagcaaga gattccagcc ctccacagcg ttccgacggg acgtgagcga ttccacgac 1680
agcgtgaggg atcctaagac cagcgagatc ctggacatca gcccttgccg cttcggcggc 1740
gtgtccgtga tcaccccggc caccacagcc agcagcgagg tgcgcgtgct gtaacagcac 1800
gtgaactgca ccgactgtgag caccggcctc cagcccgacc agctcacccc cgcctggaga 1860
atctacagca ccggcaacaa cgtgttccag acccagggccy gctgctcatt cggcgccgag 1920
cacgtggaca ccagctacga gtgcgacatc cccatcgagg ccggcatctg cgcagctac 1980
cacaccgtga gctgctgagc agcaccagc cagacagca tctgtgctca caccatgagc 2040
ctgggcgcct ga 2052

```

<210> SEQ ID NO 51

<400> SEQUENCE: 51

000

<210> SEQ ID NO 52

<400> SEQUENCE: 52

000

<210> SEQ ID NO 53

<211> LENGTH: 231

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Minimal optimization of E protein

<400> SEQUENCE: 53

```

atgtatagtt ttgtgagtga ggaagcgggc aacctgattg tcaactcagt gctgctgttc 60
ctggcctttg ttgtcttctc gctggttaact ctggcctatc tgactgccct gagactgtgc 120
gctactgctt gcaacatoyt gaactctctt ctggttaagc ccacagtta cgtgtattct 180
aggtgaaga acctgaactc cagcgagggc gtcccgatc tctgtgtatg a 231

```

<210> SEQ ID NO 54

<211> LENGTH: 1542

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Optimized soluble S2 protein with MET

-continued

<400> SEQUENCE: 54

```

atggatagtt caattgctta ctctaataac accattgcta taccactaa cttttcaatt    60
agcattacta cagaagtaat gactgtttct atggctaaaa cctccgtaga ttgtaatatg    120
tacatctgcy gagattctac tgaatgtgct aatttggctc tccaataggy tagcttttgc    180
acacaaacta atcgtgcaact ctccaggtatt gctgctgaac aggatcgcaa cacacgtgaa    240
gtgttcgctc aagtoaaaaa aatgtacaaa accccaactt tgaatattt tgggtgtttt    300
aatttttcaa aaattattcc tgaacctcta aagocaaata agaggtcttt tattgaggac    360
ttgctcttta ataaggtgac actcgtgat gctggcttca tgaagcaata tggcgaatgc    420
ctaggtgata ttaastgotag agctctcatt tbtgcgcaga agttcaatgy acttacagty    480
ttgccaactc tgcctaactga tgaatgatt gctgcttaca ctgctgctct agttagtgyt    540
actgccaacty ctggatggac atttggctct ggcgtgctc tccaatacc ttttgcatag    600
caaatggcat ataggttcaa tggcattgga gttaccacaa atgttctcta tgagaaccaa    660
aaacaaatgy ccaaccaatt taacaagcgy attagtcaaa tccaagaatc acttacaaca    720
acatcaacty cattgggcaa gctgcaagac gtgtttaacc agaattgcta agcattaaac    780
acacttgtta acaacattag ctctaatitt ggtgcaattt caagtgtgct aatgatact    840
ctttcgcgac ttgataaagt cgaggcggag gtacaattg acaggttaat tacaggcaga    900
cttcaagcgc ttcaaaccta tgaacacaaa caactaatca ggcgtgctga aatcagggct    960
ttgctcaatc ttgctgctac taaaattgtct aagtgtgttc ttggacaatc aaaaagagtt    1020
gacttttgyt gaagaggcta ccaccttatg tcttcccaac aagcagcccc gcatgtgttt    1080
gtcttctac atgtcaacta ttgcccaccc caggagagya acttaccac agcgcacaga    1140
atttgcctat aaggcaaaagc ataactccct cgtgaagtyt tttttgtgtt taatggcaat    1200
tcttggttta ttacacagag gaactctttt tctccacaaa taattactac agacaataca    1260
tttgctctag gaatttgya tgtcgttatt ggcactcata acaacacagt ttatgactct    1320
ctgcaacctg agctcgaact atccaagaaa gagctggaca agtacttcaa aaatcataca    1380
tcacacagat ttgatcttgy cgacatttca ggcattaaay cttctgtcgt caacttcaa    1440
aaagaaatgy accgcctcaa tgaagtgctc aaaaatttaa tgaatcact cattgacctt    1500
caagaattgy gaaatattga gcaatattt aaatggcctt gg                            1542

```

<210> SEQ ID NO 55

<211> LENGTH: 9

<212> TYPE: PRN

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: H-2Kd binding peptide

<400> SEQUENCE: 55

Thr Tyr Gln Arg Thr Arg Ala Leu Val

1

5

<210> SEQ ID NO 56

<211> LENGTH: 514

<212> TYPE: PRN

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Optimized S2 protein with MET

-continued

<400> SEQUENCE: 56

```

Met Asp Ser Ser Ile Ala Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr
1           5           10           15

Asn Phe Ser Ile Ser Ile Thr Thr Glu Val Met Pro Val Ser Met Ala
20           25           30

Lys Thr Ser Val Asp Cys Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu
35           40           45

Cys Ala Asn Leu Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn
50           55           60

Arg Ala Leu Ser Gly Ile Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu
65           70           75           80

Val Phe Ala Gln Val Lys Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr
85           90           95

Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro
100          105          110

Thr Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu
115          120          125

Ala Asp Ala Gly Phe Met Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile
130          135          140

Asn Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val
145          150          155          160

Leu Pro Pro Leu Leu Thr Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala
165          170          175

Leu Val Ser Gly Thr Ala Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala
180          185          190

Ala Leu Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly
195          200          205

Ile Gly Val Thr Gln Asn Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala
210          215          220

Asn Gln Phe Asn Lys Ala Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr
225          230          235          240

Thr Ser Thr Ala Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala
245          250          255

Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala
260          265          270

Ile Ser Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu
275          280          285

Ala Glu Val Gln Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu
290          295          300

Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala
305          310          315          320

Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln
325          330          335

Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe
340          345          350

Pro Gln Ala Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
355          360          365

Pro Ser Gln Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu
370          375          380

Gly Lys Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr
385          390          395          400

```

-continued

Ser Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr	
405 410 415	
Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile	
420 425 430	
Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe	
435 440 445	
Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro Asp Val	
450 455 460	
Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val Asn Ile Gln	
465 470 475 480	
Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys Asn Leu Asn Glu Ser	
485 490 495	
Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr Glu Gln Tyr Ile Lys Trp	
500 505 510	
Pro Trp	
 <210> SEQ ID NO 57	
<211> LENGTH: 1242	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Fragment of S protein	
 <400> SEQUENCE: 57	
gtgcacatgg ttatctttct gctgttcttc accctcacca ggggcagcga tctggatagg	60
tgcaccacct togacgacgt gcaaggccccc aactacaccc agcacaccag cagcatgagg	120
ggcgtgtact acccccgacga gatcttcaga agcgacaccc gtacctcacc ccaaggacctg	180
ttcctgacct tctacagcaa cgtgacccgg ttccacacca tcaaccacac cttcggaac	240
cccgatgacc ctttcaagga cggcatctac ttgcgcgcaa cagagaagag caatgtgggt	300
cggggctggg tgttcggcag caccatgaac aacaagagcc agagcgtgat cctcatcaac	360
aacagaccaa acgtgtgtat ccgggcctgc aatttcagag tgtgcacaa cctttcttc	420
gcggtgtcaa aacctatggg caccagaccc cacacatga tcttcagcaa cgccttcaac	480
tgcaccttcg agtacatagc cgaagccttc agcctggatg tgaagagaaa gagcggcaac	540
ttcaagcacc tgcgggagtt cgtgttcaag aacaaggagc gcttctgtga cgtgtcaag	600
ggctaccaga ccatcgacgt ggtgagagac ctgcaccagc gcttcaaac cctgaagccc	660
atcttcaagc tgcacctggg catcaaacct accaactccc gggcatctct caacgccttt	720
agccctgccc aggtatatctg gggcaccaag gccctgacct acttcgtggg ctacctgaag	780
cctaccacct tcatgctgaa gtacgacgag aacggcacca tcaacgatgc cgtggatgac	840
agcagaagcc ccttggcoga gctgaagtgc agcgtgaaga gcttcagatg cgaagaaggc	900
atctaccaga cagaacaact cagagtgtgt cctacggcgg atygtgtgag gttccccaat	960
atcaccaccc tgtgccctct cggcgaggtg ttcaacgcaa caagttccc tagcgtgtac	1020
gcctgggagc ggaagaagat cagcaactgc gtggcgattt acagctgtgt gtacactccc	1080
accttcttca gcaacttcaa gtgctaacgc gtgagcogca caaagctgaa cgaactgtgc	1140
ttcagcaaac gtgacgoga ctacttctgt gtgaaggcgg acgacgtgag acagatcgcc	1200
cctggccaga ccggcgtgat cgcgaactac aactacaaga tt	1242

-continued

```

<210> SEQ ID NO 58
<211> LENGTH: 412
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Fragment of S protein

<400> SEQUENCE: 58
Met Val Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
1           5           10           15
Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
20           25           30
His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
35           40           45
Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
50           55           60
Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
65           70           75
Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
85           90           95
Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
100          105          110
Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
115          120          125
Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
130          135          140
Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
145          150          155          160
Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
165          170          175
Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly
180          185          190
Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
195          200          205
Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu
210          215          220
Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro
225          230          235          240
Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Tyr Phe Val Gly Tyr
245          250          255
Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile
260          265          270
Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys
275          280          285
Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn
290          295          300
Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr
305          310          315          320
Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser
325          330          335
Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr
340          345          350

```

-continued

Ser	Val	Leu	Tyr	Asn	Ser	Thr	Phe	Phe	Ser	Thr	Phe	Lys	Cys	Tyr	Gly
	355						360					365			
Val	Ser	Ala	Thr	Lys	Leu	Asn	Asp	Leu	Cys	Phe	Ser	Asn	Val	Tyr	Ala
	370				375					380					
Asp	Ser	Phe	Val	Val	Lys	Gly	Asp	Asp	Val	Arg	Gln	Ile	Ala	Pro	Gly
	385				390				395					400	
Gln	Thr	Gly	Ile	Ile	Ala	Asp	Tyr	Asn	Tyr	Lys	Leu				
			405						410						

<210> SEQ ID NO 59
 <211> LENGTH: 1432
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Fragment of S protein
 <400> SEQUENCE: 59

```

aagcttcccg acgaacttoat gggctgctg ctggcctgga acaccagaaa catcgacgcc      60
acctccaccg gcaactacaa ttacaagtac cgtacactga ggcacggcaa gctgagaccc      120
ttcgagcggg acatctccaa cgtgccttc agccccagag gcaagccctg caccocccct      180
gcccgaact gctactggcc cctgaacgac tacggctctc acacacacac cggcatcgcc      240
tatcagccct acagagtggt ggtgctgagc ttcgagctgc tgaacgcccc tgcacccgtg      300
tggcgcccca sgctgagcac cgacctcacc aagaaccagt gctgaactt caactcaacc      360
ggcctccacc gcaaccgctt gctcaccccc agcagcaaga gattccagcc cttccagcag      420
ttcggcaggg acgtgagcga ttccaccgac agcgtgaggg atcctaaagac cagcgagatc      480
ctggacatca gcccttgca gcttcggcgg gtgtccgtga tcaaccccg caccacagcc      540
agcagcgagg tggcgctgct gtaccaggac gtgaactgca ccgacgtgag caccgcacc      600
caccgcgacc agctcaccoc cgcttgagga attctacaga ccggcaacaa cgtgttccag      660
accacggccg gctgcctcat cggcgccgag cacytgga ccaactacga gtgcgacatc      720
ccatcggag cggcagctct gcacagctac cacacgtga gctgtgtag aagcaccaga      780
cagaagagca tcgtggccta caccatgagc ctggggcgcc acagcagcat cgcctacagc      840
aacaacacca tcgcctaccc caccacattc agcatctcca taccaccca ggtgatgcc      900
gtgagcatgg ccaagaccag cgtgattgc aacatgtaca tctcgggcga cagcaccggc      960
tgcgccaaac tctgctgtag gtacggcagc ttctgcaccc agctgaacag agccctgagc      1020
ggcattgccg ccgagcagga cagaacacac agggaggtgt tcgcccaggt gaagcagatg      1080
tataagacc ccacccctgaa gtacttcggc gggttcaact tcaagccagat cctgcgccat      1140
cctctgaagc ccaccaagcg gacgttcacc gaggacctgc tgttcaacaa ggtgaacctg      1200
gccgacgccg gctttatgaa gcagtacggc gagtgcctgg gcgatatcaa cgccaggagc      1260
ctcatctcgc ccgagaagtt caacggcttg acctgcctgc cccctctgct caccgatgat      1320
atgatcgccg cctatacagc cgcctcgttg tcaggacccg ccaccgccgg ctggaccttt      1380
ggcgccggag ccgccttgaa gatcccttc gccatcgaga tggcctaccg gt      1432
  
```

<210> SEQ ID NO 60
 <211> LENGTH: 1118
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:

-continued

<223> OTHER INFORMATION: Fragment of 5 protein

<400> SEQUENCE: 60

```

accggttcaa tggcatcggc gtgacccaga acgtgtgtga cgaagaaccag aagcagatcg      60
ccaaaccagt caataaggcc atctcccaga tccagggagc cctcaacacc acaagacacg      120
ccctgggcaa gctgcaggac gtgtgtaacc agaagccca ggcctgaat accctggtga      180
agcagctgag cagcaacttc ggcgcacatc gaacgtgtgt gaacgacatc ctgagcaggc      240
tggataaggt ggaaggccag gtgcagatcg acagactcat caccggcaga ctgcagagcc      300
tgcagaccta cgtgaaccag cagctcatca gagccgcaga gatcagagcc agcgcaaatc      360
tggcgcgccac caagatgagc gagtgtgtgc tgggcagagc caagagagtg gaattctgctg      420
gcaagggtta tcaactcatg agttccctc aggcgcgtcc caacgggctg gtgttctctg      480
acgtgaccta cgtgcctagc caggagagga atttaccacc gcgccacagc atctgcacag      540
agggcaaggg ctacttcccc agagaggcgc tgttctgttt taacggcacc agctggttca      600
tcacccagcg gaacttcttc agccccaga tcataccacc agacacaccc ttctgttccg      660
gcaattgaga cgttgctcat ggcatactca ataaccacct gtacgacccc ctgcagcccg      720
agctggatag cttcaaggag gagctggaca agtaactcaa gaacacaccc tccccgacg      780
tggacctggg cgacatcagc ggcataatg ccagcgtggt gaacatccag aaggagatcg      840
accggctgaa cgaaggtgcc aagaaactga acgagagcct catcgacctg cagcagctgg      900
gaaagtacga gcagtacatc aagtggccct ggtacgtgtg gctgggtctc atcgccggcc      960
tcactggcaat cgtgatggtg accatcttgc tgtgctgcat gaccagctgc tgcctctgcc      1020
tgaagggcgc ctgcagctgt ggcagctgct gcaagttcga cgaagacgac tcagagcccg      1080
tgtgaaaggg cgtgaagctg cactacacct gaagatct      1118

```

<210> SEQ ID NO 61

<211> LENGTH: 3780

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Mutated 8 protein

<400> SEQUENCE: 61

```

gtgacatgg ttatctttct gctgtcttc accctaccca gcygcagcga tctggatagg      60
tgcacacct tcgaacagct gcaagcnccc aactacaccc agcacacccag cagcatgagg      120
ggcgtgtact accccgacga gattttcaga agcagacccc tgtacctcac ccaggacotg      180
ttcctgacct tctacagcaa cgtgacggcg ttccacacca tcaaccacac cttcggcaac      240
ccctgtatcc ctttcacgga cgcacatcac ttccgcgcga ccgagaaagag caattgtggt      300
cgggcgtggg tcttcggcag caccatgaac aacaagagcc agagcgtgat catcatcaac      360
aacagcccca accgtgtgat ccggcgtcgc aatttcagag tgtgcagaaa ccctttcttc      420
gcccgttcca accctatggg caccacagacc cacacatgta tcttgacaaa gcgcttcaac      480
tgcaccttgg agtcatcagc cgacgccttc agcctggatg tgagcgagaa gagcgcgaac      540
ttcaagcacc tgcgggagtt cgtgttcaag acaagggacg gcttcttgta cgtgtacaa      600
ggctaccagg ccatcgagct ggtgagagac ctgcaccagc gcttcaaac cctgaagccc      660
attctcaagc tgcacctggg catcaacatc acaaacttcc gggccatcct caccgccttt      720

```

-continued

agccctgccc agyatatctg gggcaccagc gccgctgctt acttcgtggg ctacctgaag	780
ctaccacct tcatgctgaa gtaacagag aaaggaacca tcaacgatgc cgtggactgc	840
agccagaacc ccttgccgca gctgaagtg agcgtgaaga gcttcgagat cgaacaagggc	900
atctaccaga ccaagcaactt cagagtggtg cctagcggcg atgtggtgag gttcccacat	960
atcaaccaac tgtgcccttt cggcgaggtg ttcaacgcca ccaagttccc tagcgtgtac	1020
gcctgggagc ggaagaagat cagcaactgc gtgcccattt acagcgtgct gtacaactcc	1080
accttcttca gcaaccttcaa gtgtacggc gtgagcgcca ccaagctgaa cgacctgtgc	1140
ttcagcaacg tgtacgccga ctcatctgtg gtgaaggcgg acgacctgag acagatcgcc	1200
cctggccaga ccggcgtgat cgcgcactac aactacaagc ttcccgacga cttcatgggc	1260
tgcgtgctgg cctggaaac cagaacatc gacgccacct ccacggcga ctacaattac	1320
aagtaccgct acctgaggca cggcaagctg agaccccttg agcgggacat ctccaacgtg	1380
cccttcagcc ccgacggcaa gccctgcacc cccctggccc tgaactgta ctggcccttg	1440
aacgactacg gctctacac caccaccgga atcggtatc agccctacag agtggtgtgtg	1500
ctgagtttcg agctgctgaa cgcacctgcc accgtgtgct gccccaagct gagcaacgac	1560
ctcatcaaga accagtgctg gaacttcaac ttcaacggcc tcaacggcac cggcgtgctc	1620
acccccagca gcaagagatt ccagcccttc cagcagtttg gcaggagcgt gagcgatttc	1680
accgacagcg tgaggatccc taagaccagc gagatcctgg acatcagccc ttgcagcttc	1740
ggcggctgt cctgtgtcac ccccgccacc aacgccaga gcgaggtggc cgtgcgtgac	1800
caggagctga actgcaccga cgtgagcaac gccatccacg ccgaccagct caaccccgc	1860
tgagagaact acagcacagg caacaacgtg ttccagaccc agccggctg cctcatcggc	1920
gcgaggaacg tggacaccag ctacgagtg gacatcccca tcggagccgg catctgcgc	1980
agctaccaca cgttgagcct gctgagaagc accagccaga agagcatcgt ggcctacacc	2040
atgagcctgg gcgcgacag cagcatcgcc tacagaacca acacatcgc catcccacc	2100
aacttcayca tctccatcac caccgagtg atgcocctga gcatggccaa gaccagcgtg	2160
gattgcaaca tgtacatctg cggcgacaga accgagtgcc ccaacctgct gctgcagtac	2220
ggcagcttct gcacccagct gaacagagcg ctgagcggca ttgcgccga gcaggacaga	2280
aacaccaggg agsgtttgc ccaggtgaag cagatgtata agaccccac cctgaagtac	2340
ttcggcgggt tcaactcaag ccagatcctg cccgactctc tgaagccac caacggcagc	2400
ttcatcgagg aactcgtgtt caacaagtg accctggccg acgcggctt tatgaacag	2460
tacggcgagt gccctggcga tatcaacgac agggactcta tctgcgcca gaagtcaac	2520
ggcttgaccg tgcctccccc tctgctcacc gatgatatga tcgcgccta tacagccgc	2580
ctggtgtcag gcaccgccac cgcggctcgg accttggcg ccggagccgc cctgcagata	2640
cccttcgcca tgcagatgac ctaccggttc aatggcatcg gcgtgaacca gaactgctg	2700
tacgagaacc agaagcagat cgcacaaccg ttcaataagg ccatctacca gatccaggag	2760
agctcaccac ccaacaagac cgcctgggag aagctgcaag acgtgtgtaa ccaagacgcc	2820
caggccctga atacctggtt gaagcagctg agcagaacct tcggcccat cagcagcgtg	2880
ctgaagacaa tctgagcag gctggataag gtgagggcgg agtgtcagat cgacagactc	2940
atcaacggca gactgcagag cctgcagacc tacttgaccc agcagctcat cagagccgc	3000

-continued

```

gagatcagag ccagcgccaa tctggccgc accaagatga gagatgcyt gctgggcccag 3060
agcaagagag tggacttcgt cggcaagggc tatacaatca tgagottccc tcaggccgct 3120
ccccacggcg tgggtttcct gacgtgaac tactgtccta gccaggagag gaattcaacc 3180
accgccccag ccatctgcca cgaaggccaag gctacttccc ccagagaggg cgtgttcgtg 3240
tttaacggca ccagctgggt catcccccag cggaaattct tcagcccccga gatcatcacc 3300
acagacaaca ccttcgtgtc cggcaattgc gacgtggcca tcggcatcat caataacacc 3360
gtgtacgaac ccttcagacc cgaactggat agcttcaag aggagctgga caactacttc 3420
aagaaccaca cctccccga cgtggacctg ggcgacatca gcggcatcaa tgcacgctg 3480
gtgascaccc agaaggagat cgcacggctg aacgaggtgg ccaagaacct gancgagagc 3540
ctcatcgacc tgcaggagct gggaaagtac gacgcatac tcaagtggcc ctggtacgtg 3600
tggctgggct tcatcgccgg cctcatcgcc atcgtgatgg tgaccatcct gctgtgctgc 3660
atgaccagct gctgctcctg cctgaagggc gctgcaagct gtggcgctg ctgcaagttc 3720
gacgaggagc actcagagcc cgtgctgaag ggcgtgaagc tgcactacac ctgaagatct 3780

```

<210> SEQ ID NO 62

<211> LENGTH: 1255

<212> TYPE: PRN

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Mutated S protein

<400> SEQUENCE: 62

```

Met Val Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
1           5           10          15
Asp Arg Cys Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
20          25          30
His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
35          40          45
Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
50          55          60
Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
65          70          75
Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
85          90          95
Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
100         105         110
Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
115         120         125
Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
130         135         140
Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
145         150         155
Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
165         170         175
Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe His Asn Lys Asp Gly
180         185         190
Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
195         200         205
Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu

```


-continued

210	215	220
Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro 225 230 235 240		
Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Tyr Phe Val Gly Tyr 245 250 255		
Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile 260 265 270		
Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys 275 280 285		
Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn 290 295 300		
Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr 305 310 315 320		
Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser 325 330 335		
Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr 340 345 350		
Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly 355 360 365		
Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala 370 375 380		
Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly 385 390 395 400		
Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe 405 410 415		
Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser 420 425 430		
Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu 435 440 445		
Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly 450 455 460		
Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp 465 470 475 480		
Tyr Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val 485 490 495		
Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly 500 505 510		
Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn 515 520 525		
Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg 530 535 540		
Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp 545 550 555 560		
Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys 565 570 575		
Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser 580 585 590		
Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr 595 600 605		
Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr 610 615 620		

-continued

Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu		
625	630	635 640
His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile		
645	650	655
Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys		
660	665	670
Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala		
675	680	685
Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile		
690	695	700
Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys		
705	710	715 720
Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu		
725	730	735
Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile		
740	745	750
Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys		
755	760	765
Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe		
770	775	780
Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile		
785	790	795 800
Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met		
805	810	815
Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile		
820	825	830
Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Leu Leu Thr		
835	840	845
Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala		
850	855	860
Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe		
865	870	875 880
Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn		
885	890	895
Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala		
900	905	910
Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Ser Thr Ala Leu Gly		
915	920	925
Lys Leu Gln Asp Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu		
930	935	940
Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn		
945	950	955 960
Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp		
965	970	975
Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln		
980	985	990
Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala		
995	1000	1005
Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp		
1010	1015	1020

-continued

Phe Cys	Gly Lys	Gly Tyr	His	Leu Met	Ser Phe	Pro	Gln Ala	Ala
1025			1030				1035	
Pro His	Gly Val	Val Phe	Leu	His Val	Thr Tyr	Val	Pro Ser	Gln
1040			1045			1050		
Glu Arg	Asn Phe	Thr Thr	Ala	Pro Ala	Ile Cys	His	Glu Gly	Lys
1055			1060			1065		
Ala Tyr	Phe Pro	Arg Gly	Gly	Val Phe	Val Phe	Asn	Gly Thr	Ser
1070			1075			1080		
Trp Phe	Ile Thr	Gln Arg	Asn	Phe Phe	Ser Pro	Gln	Ile Ile	Thr
1085			1090			1095		
Thr Asp	Asn Thr	Phe Val	Ser	Gly Asn	Cys Asp	Val	Val Ile	Gly
1100			1105			1110		
Ile Ile	Asn Asn	Thr Val	Tyr	Asp Pro	Leu Gln	Pro	Glu Leu	Asp
1115			1120			1125		
Ser Phe	Lys Glu	Glu Leu	Asp	Lys Tyr	Phe Lys	Asn	His Thr	Ser
1130			1135			1140		
Pro Asp	Val Asp	Leu Gly	Asp	Ile Ser	Gly Ile	Asn	Ala Ser	Val
1145			1150			1155		
Val Asn	Ile Gln	Lys Glu	Ile	Asp Arg	Leu Asn	Glu	Val Ala	Lys
1160			1165			1170		
Asn Leu	Asn Glu	Ser Leu	Ile	Asp Leu	Gln Glu	Leu	Gly Lys	Tyr
1175			1180			1185		
Glu Gln	Tyr Ile	Lys Trp	Pro	Trp Tyr	Val Trp	Leu	Gly Phe	Ile
1190			1195			1200		
Ala Gly	Leu Ile	Ala Ile	Val	Met Val	Thr Ile	Leu	Leu Cys	Cys
1205			1210			1215		
Met Thr	Ser Cys	Cys Ser	Cys	Leu Lys	Gly Ala	Cys	Ser Cys	Gly
1220			1225			1230		
Ser Cys	Cys Lys	Phe Asp	Glu	Asp Asp	Ser Glu	Pro	Val Leu	Lys
1235			1240			1245		
Gly Val	Lys Leu	His Tyr	Thr					
1250			1255					

<210> SEQ ID NO 63

<211> LENGTH: 1281

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Mutated N protein

<400> SEQUENCE: 63

gtcgacatga gcgacaacgg cccccagagc aaccagagaa ggcgccccag aatcaccttt	60
ggcgggcccta ccgacagacac cgacaacaac cagaacggcg gcagaaaacg gcgcagaccc	120
aagcagagga gacccccagg cctgcconac aacacgcgca gctggttcac cgccctcacc	180
cagcacggca aggaggagct gagattcccc agaggcaggg gctggcccat caatacaca	240
agcgggcccc agcatcagat cggctactac cggagggcga ccagaagaat gagaggcggc	300
gacggcaaga tgaaggagct gagccccgg tggtaattct actactggg caccggccct	360
gaggcgaccc tgcctacagg cgcacaacag gagggcatcg tgtgggtggc caccggaggg	420
gccttgata cccccagga ccaatcggc aacaggaaac caacaacaa tgcgcgaacc	480
gtgctgcagc tgccccaggg caccacctg cccaagggct ttacgcgga gggcagcaga	540

-continued

```

ggcggcagcc aggcacagac cagaagcagc agcaggagca ggggcaacag cagaatagc 600
accccccggca gcagcagagg aaattacccc gccagaaatg ccaagggggg agggcgagac 660
ggccttgccc tgcgtctctt ggcagagctg aatcagcttg agagcaaggt gagcgcaag 720
ggccagcaac agcaggagca gaccttgacc aagaagtctg ccgcccaggc cagaagaag 780
ccccggcaga agagacccgc caccacgagc tacaaatgtg cccagcctt cggcagaga 840
ggccccggcgc agaccacagg caatttcggc gacnaggacc tcatacagca gggcaccgac 900
tacaagcact ggctcagat cgcacagttc gcccccagcg ccagcgctt cttcgccatg 960
agccggatcg gcatggaggt gacccccagc ggcacctgac taactacca cggcgccatc 1020
aagctggagc acaaggaccc ccagttcaag gacaaagtga tctctgtgaa caagcacatc 1080
gacgctaca agaccttccc acccaccgag cccaaagagg caaagaagaa gaacaccgac 1140
gagggccagc cctgcctcca gagacagaag aagcagccca ccgtgacct gctgcttgc 1200
gcgcacatgg acgacttcag ccgcacagtg cagaatagca tgagcgggcg ccttgccgat 1260
tcaccccagg cctgaagatc t 1281

```

<210> SEQ ID NO 64

<211> LENGTH: 1542

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Uniform optimization of S2 protein with MET

<400> SEQUENCE: 64

```

atggacagca gcatcgcta cagcaaacac accatcgcca tccccaccaa cttoagcatc 60
agcatcacca ccgaggtgat gccctgagc atgcacaga ccaagctgga ctgcacatg 120
tacctctgag gcacacagac agagtggccc aaactgtctg tgcagtacgg cagctctctg 180
accacgtga accgggcccct gagcggcatc gccgcaggac aggaccggaa caccggggag 240
gtgttgcccc aggtgaagca gatgtacaag acccccaccc tgaagtactt cggcggttc 300
aaetccagcc agatctgccc cgaccccctg aagccaccca agcgagctt catcgaggac 360
ctgctgttca acagagtgac cctggcgac gccggcttta tgaagcagta cggcgagtg 420
ctggggcaca tcacagcccg ggaactgata tgcgccagca agttcaacgg cctgacgtg 480
ctgcaccccc tgcataccga cgcactgata gccgcctaca ccgcccgcct ggtgagcggc 540
accgcacagg ccgctgtgac cttggcgccc ggcgcggccc tgcagatccc cttcgccatg 600
cagatggcct acccggttaa cgcactcggc gtgacccaga acgtgtgtga cagaacacag 660
aagcagatcg ccaacaaagt caacaaaggcc atcagccaga tccaggagag cctgaccacc 720
accagcaccg ccttgggcaa cgtgcaggag gtgtggaacc agaacgccca ggcctgaac 780
accctggtga agcagctgag cagcaacttc ggcgcctaca gcagcgtgct gaacgacatc 840
ctgagccggc tggaacaagt ggaaggccag gtgcagatcg accggtgat caccggccgg 900
ctgcagagcc tcagaaacta cgtgacccag cagctgatcc gggccgcgga gatccgggcc 960
agccgcaacc tggccggcac caagatgagc gagtctgtgc tggccagagc caaggggtg 1020
gacttctgag gcaagggtta ccaactgatg agcttcccc aggcggccccc ccaaggggtg 1080
gtgttctgag cgtgaccta cgtgccacga caggagcgga acttaaacac cgcacccggc 1140
atctgccacg agggcaaggc ctacttcccc cgggagggcg tgttgtgtt caacggcacc 1200

```

-continued

agctggttca	tcacccacag	gaactttctc	agcccccaga	tcataccac	cgacaacac	1260
ttcgtgagcg	gcaactgcga	cgtgtgtgac	ggcatcatca	acaaacogt	gtacgaaccc	1320
ctgcagcccg	agctggacag	cttcaaggag	gagctggaca	agtacttcaa	gaacacacac	1380
agcccccagc	tggacctggg	cgacatcagc	ggcatcaacg	ccagctgggt	gaacatccag	1440
aaggagctcg	accggctgaa	cgaagtgccc	aagaacctga	acgagagcct	gatcgacctt	1500
caggagctgg	gcaagtacga	gcagtacatc	aagtgccctt	gg		1542

<210> SEQ ID NO 65

<211> LENGTH: 1542

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Fully optimized S2 protein with MET

<400> SEQUENCE: 65

atggacagtt	caatgoccta	ttcgaaacaa	actatagcaa	tcocaaacaaa	tttttcaatt	60
tcataacaaa	cagaggtgat	gccagtgctc	atgcacaaaga	ctagcgtaga	ctgcaatatg	120
tacatctgcg	gagattctac	agaattgtca	aaactgtctg	tacagtatgg	atcgttctgt	180
acccagctca	acggggcgct	gagcggcatt	gctgcgaac	aggatcgcaa	taccagagag	240
gtgtttgtct	aagtgaacaa	aatgtataag	accocaaact	tgaataactt	cgttgatctc	300
aatttcagtc	agattctctg	agacccactc	aaacccaccca	agagagcttt	tattgaagat	360
ctttctgtca	acaaagttac	cttggccgac	gctgggttta	tgaagcaata	cgttgagtcg	420
ctgggocaga	ttaacgcacg	agacctgata	tgcgcocaga	agtttaacgg	gctcagcggt	480
ttacgcacac	tgcctgactga	tgatattgatt	gocgcttaaa	ctgcggccct	tgtgagtggt	540
acgcgaactg	ctggctggac	gtttggcgct	ggggcgccct	tacagatccc	ttttgocatt	600
cagatggcct	acaggttcaa	tgaatttggt	gtcactcaga	atgtctctgt	cgagacacag	660
aaacagatcg	ccaaacagtt	caataaagct	atttcacaga	ttcaggaata	acttaccaca	720
acttcacagg	cactcgttaa	actgcagaa	gtgtgaaatc	agaagctcaa	ggcactaaat	780
acactcgtca	agcaactgag	ttccaattto	gyggccatatt	ctagcgtatt	gaacyacata	840
ctcagtcgca	tgcacaaagt	ggagggccga	gtccaaatag	accgtcttat	cacaggcaga	900
ctacagtcatt	tgcagaccta	cgttaacccg	cagttgatcc	gcgcgcgtga	gatacagccc	960
tcgcgcacac	tggcgcgtac	caaatgtctc	gagtgctgca	tcggacaaag	taacggggtg	1020
gattttttgg	gaaagggcta	tcacatcag	tccttccctc	aagcagcacc	ccacggagtc	1080
gtttttctgc	atgtgacata	cgtgctagca	caggagagaa	actttaccac	tgcgcctgac	1140
atttgtcatg	aaggcaaaag	ttattttccc	cgcgaggggg	tgttctgttt	caacggaaat	1200
agctggttta	tcacacaaag	gaatttcttc	tcocccocaga	tcataccac	cgacaacac	1260
ttgtctctcg	gaactcttga	cgtctgtata	ggcatcatca	ataatacagt	atacgtatcc	1320
ctgcagcccg	aacttgactc	tttcaaggag	gaactagata	agtacttcaa	gaatcacacc	1380
agcccgagtg	tagatttagg	ggatatttag	gggattaaag	catcgtgggt	caacatccaa	1440
aaagagattg	acagactgaa	cgaagtggcg	aagaacctga	atgagtcctt	gatcgacctt	1500
caggagctgg	gcaagtatga	acagtatatc	aagtgccctt	gg		1542

<210> SEQ ID NO 66

-continued

```

<21> LENGTH: 1542
<22> TYPE: DNA
<23> ORGANISM: Artificial Sequence
<20> FEATURE:
<23> OTHER INFORMATION: Minimal optimization of S2 protein with MET
<400> SEQUENCE: 66

atggatagca gcatagccta ctcaaacac acgatogcaa tccccacaa cttttccatt 60
tcataaacta ccgaggtgat gccctgtagc atggccaaag catcgttaga ttgcaacatg 120
taaatctgtg cggattctac agagtgtgcc aacctgctgc tgcagtcagg ctctttctgc 180
acgcagctga acagggccct gtctggcctc ccgcgcagac aggatcgcaa cacacgggag 240
gttttcgcc cgytaaagca gatgtataag acgccacctc tgaagtactt cggcggcttc 300
aaattctctc agatactgac cgacccctgc aagccacta agagtcttt tatcgaggat 360
ctgctgttca acaaggttac cctggccgat gccgcttta tgaagcagta tggcagagtgc 420
ctggggcaca tcaacggccg agatctgata tgcgccaga agttcacagg cctgactgtg 480
ctgcaccccc tgcgtactga cgcacatgac gccgcctata ccgcgcacct ggtgagtgcc 540
acagccactg ccgctggac attcggcgcc ggccgcgcgc tgcagatccc cttcgccatt 600
cagatggcct acagatttaa ccgcatgggc gtcaactcaga acgtcctgta tgagaccag 660
aagcagatcg ccaaccagtt taacaaagcc ataacgcaga tccaggagtc actgacaaag 720
acaagtacgg cctcgggcaa gctgcagagt gtatggaacc agaaccgcca ggccctgaac 780
actctgttta agcagctgtc tagcaacttc ggccgcatca gtatgtttct gaacgatatt 840
ctgtctagtc tggacaaggt cgaagccgag gtgcagattg atcgctgat taaccggcga 900
ctgcagagtc tgcagactta tgaactcag cagctgatac gagccgcga gattcgagcc 960
tcgcgcaccc tggccgcacc aaagatgtct gagtgcgtcc tgggcacag taagaggggt 1020
gaattctcgg gcaaggccta tcatctgatg tcttttcccc agycgcgcgc caacggcgct 1080
gtgttctcgc acytaactta cgtgccaggt caggagagaa acittaccac tgcgcccgcc 1140
atctgcacag aggggcgaagc ctacttcccc agagagggcg tgtttgtgt caacggcaca 1200
tcttggttca tcaaccagag gaacttttcc agccccaga tcaaacacac tgcaaacact 1260
ttcgtttcgg gcaactcgca cgtagtgtac ggcataataa acaaacccgt gtaacatccc 1320
ctgcagcccg agctggacac ctttaagag gagctggaca agtaacttaa gaacataacc 1380
tcacccgatg tggacctggg cgacatttct ggcataaacy cctcctcgt caacatccag 1440
aaggagatag atagactgaa cgaggttgcc aagaacctga acgagtcctg gatcgactcg 1500
caggagctgg gcaagtacga gcagtatata aagtggccct gg 1542

<210> SEQ ID NO 67
<21> LENGTH: 1588
<22> TYPE: DNA
<23> ORGANISM: Artificial Sequence
<20> FEATURE:
<23> OTHER INFORMATION: Standardized optimization of soluble S protein
<400> SEQUENCE: 67

atgtcaatct tctgtctgtt cctgacctg accagcgcca gcaactgga tcgtcgacc 60
aacctcgatg acgtgcagga ccccaactac acccagcata ccagcagcat gcgcggcggt 120
taactacccc atgagatctt ccgcagagac accctgtacc tgacccagga cctgttctgt 180

```

-continued

ccctcttaca	gcacgtgac	cggtctccac	acactcaacc	ataccttcgg	caaccccgty	240
atccccctta	aggaaggcat	ctactctggc	gccacggaga	agagcaacgt	ggtgcgcggt	300
tggtgtgttcg	gcagcaccat	gaacaacaag	agccagagcg	tgatctcat	caacaacagc	360
acacacgttg	tgatccgcgc	ctgcacattc	gagctgtgcy	acaacccctt	cttcgcgcgt	420
agcaagccca	tgggcaccca	gaacccatcc	atgatcttgg	ataacgcctt	caactgcacc	480
ttcgagtaca	tcagcagcgc	cttcagcctg	gacgtgagcg	agaagagcgg	caacttcaag	540
catctgcgcg	agttcgttgt	caagaacaag	gatggcttcc	tgtacgtgta	caagggtcac	600
cagcccatcg	cagtggtgcy	cgtactgtccc	agcggcttca	acacccctga	gcccatcttc	660
aagctgcccc	tgggcataca	catcaacaac	ttccgcgcga	tcttgaccgc	cttcagcccc	720
gcccaaggaca	tctggggcac	cagcgccgcg	gcctacttcg	tgggtacct	gaagccacc	780
accttcctgc	tgaagtaoya	tgaagaacgg	accatcacgc	acgccttgga	ctgcagccag	840
aacccccctg	ccgagctgaa	gtgcagcgtg	aagagcttcg	agatogataa	gggcactcac	900
cagaccagca	acttcccgct	ggtgcccgag	ggcagcgtgg	tgccttcccc	caacatcacc	960
aaactgtgtc	cttcggcgga	ggtgttcaac	gccacaagt	tcccccagct	gtacccctgg	1020
gagcgcaga	agatacgcaa	ctgcgtgcyg	gactacagcg	tgtgttaca	cagcaacttc	1080
ttcagcacc	tcaagtgtat	cggcgtgagc	gccaccaagc	tgaacgatct	gtgtctcagc	1140
aacgtgtacg	ccgacagctt	cgtgttgaa	ggcagtgatg	tgcgcagat	cgcgcccgcc	1200
cagaccggcg	tgcgtgcgga	ttacacatcc	aagctgcccc	acgaactcat	gggtcgcgtg	1260
ctggcctgga	acacccgcga	catcgacgcg	accagcacgc	gcaactacaa	ctacaagtac	1320
cgtacctgc	gcactggcaa	gctgcgcgcc	ttcgagcgcg	atatcagcaa	cgtgccttcc	1380
agccccgatg	gaaagccctg	caaccccccc	gcctgaact	gttactggcc	cctgaacgac	1440
tacggttct	acacacccac	cggccttcgc	taccagccct	accgcgtgg	ggtgtgagc	1500
ttcgagctgc	tgaacgcgcc	cgcacccgtg	tycggcccca	agctgagcac	gcaactgtac	1560
aagaacccgt	gcgtgaact	caacttccac	ggcctgacgc	gcacccggct	gtgaaccccc	1620
agcagcaagc	gcttccagcc	cttccagcag	ttcggccgcg	atgtgagoga	cttccacgat	1680
agcgtgcgcg	accccaagac	cagcagagtc	cttgatatca	gccccctgca	cttcggcgcc	1740
gtgagcgtga	tcacccccgg	caccacagcc	agcagagagc	tggccgtgct	gtaccaggat	1800
gtgaactgta	cagatgtgag	cacccgcctc	cacgcagatc	agctgacccc	cgcctggcgc	1860
atctacagca	ccggcaacaa	cgtgttcagc	accagggccg	gctgctgat	cggcgccagc	1920
catgttgaca	ccagctacga	gtgtgacatc	cccatggcgc	cggcactctg	tgcagcttac	1980
caacacgtga	gcctgcgcgc	cagcaacagc	cagaagagca	tctgtgccta	caactagcgc	2040
ctggcgccgc	atagcagcat	cgcctaacgc	aaacaaccca	tgcacatccc	caaccaattc	2100
agcatcagca	tcaccaacga	ggtgatgccc	gttagcatgg	ccaaagacag	cgtggactgc	2160
aaactgtaca	tctgcggcga	tacacagcag	tycgcacacc	tgtgtgtgca	ctacggcagc	2220
ttctgacccc	agctgaancc	cgccttgagc	ggcatgcgcg	ccgagcagga	tcgcaaacac	2280
cgcgaggtgt	tgcgccaggt	gaagcagatg	tacaagaccc	ccacccctga	gtactctggc	2340
ggcttcaact	tcagccagat	ctcgcgcgat	cccttgagc	ccacacagcg	cagcttcatc	2400
gaggatctgc	tgctcaacaa	ggtgacccgt	gcagatgcgc	gcttcatgaa	gcagtagccc	2460

-continued

```

gagtgccctg cgcgatccaa cgcgcgcgat ctgatctcgc ccagaaagt caacggcctg 2520
acgtgctgc ccccccctgc gaccagacac atgatcgcgc cctaacccgc cgcctgggtg 2580
agcggcacgc ccacgcgcgc ctggacatto ggcgcgcgcg ccgccttgca gatcccttc 2640
gcctgcaga tggcctacgc cttcaacggc atcggcgtga ccagaaagt gctgacagag 2700
aacagaaagc agatcgccaa ccagttcaac aagccatca gccagatcaa ggcagcctg 2760
acaaacccca gcaacgcgcct ggcgaagctg cagcagctgg tgaacagaa cgcacaggcc 2820
ctgaaccccc tggtagaaga cgtgagcaga aaattcggcg ccatcagcag cgtgctgaac 2880
gacatcctga gccgcctgga taaggtggag gccagcgtgc agatcgatgc cctgataacc 2940
ggcgcgcctg agagccttga gacctaactg accacgcaga tgatccgcgc cgcgcagata 3000
cgcgcacagc ccaacctggc cgcacacaa agtagcagtg gctgctggg ccagagcaag 3060
cgctggtatt tctgcgcgaa gggctaacaa ctgatgagct tccccacag cgcgccacct 3120
ggctggtgtt tctgcgcagt gacctaactg ccaacgcagg agcgcaactt caacacgcgc 3180
ccgcgcctct gccacggagg caaggcctac ttcccgcgcg agggcgtgtt cgtgtcaac 3240
ggcacacgct ggttcataca ccacgcgaac ttcttcagcc cccagatcat caacacgcat 3300
aacaaccttg tagcgcgcga ctgcgatttg gtgatcgcaa tcaatacaaa caactgtgac 3360
gatccccctg agccgcagct ggcagacatto aagggggaga tggataagta cttcaagaac 3420
caaacacgac ccagcgtgga tctgggcgat atcagcgcaa tcaacgcaga cgtggtgaac 3480
atccagaaag agatcgatgc cctgaacagc gtggccaaag aactgaacga gagcctgata 3540
gaactgcagg agctgggcga gtaacgcgag taatacaagt ggcctctg 3588

```

<210> SEQ ID NO 68

<211> LENGTH: 2049

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Standardized optimization of soluble S1 protein

<400> SEQUENCE: 68

```

atgttcattt tctcgtgttt cctgacctg accagcgcaa gcatctgga ccgtgcacc 60
accttcgacg atgtgcaggc ccccaactac acccagacaa ccagcagcat gcgcggcgtg 120
taataccccg atgagctatt ccgcagcagt aacctgtacc tgaacacgga tctgttctgt 180
ccctttacaa gcaactgtae cggcttccat accatacaac acaacttcgg caaccccggt 240
atccccctta aggatggcat ctacttcgcg gccaccgaga agagcaacgt ggtgcgcgcg 300
tgggtgtttg cgagcaccat gaacaaacag agccagcagc tgatcatcat caacacacgc 360
accaacgtgg tgatccgcgc ctgcacattc gagctgtgcg caaacccctt cttgcgcgtg 420
agcaacccca tgggcaccca gaaccaacac atgatcttcg acaacgcctt caactgcacc 480
ttcagataca tcagcagatc cttcagcctg gacgtgagcg agagagcgcg caacttcaag 540
catctgcgcg agttcgtgtt caagacaagc gatgcttcct tgtaactgta caagggctac 600
cagcccatcg acgtgtgtcg cgaactgcac agcggcttca caacctgaa gccatcttc 660
aagctgcctc tgggcataaa catcaccaac ttccgcgcga cctgacgcgc cttcagcccc 720
gccacgata tctggggcac cagcgcgcgc gctactcttg tggcgtacct gaagcccaac 780
accttcacgc tgaagtaaga cgaagaacgc accatcacgc atgcctgga ttgcagcag 840

```


-continued

aacccccctgg ccgagctgaa gtgcagcgtg aagagcttcg agatcgataa gggcatctac	900
cagaccagca acttcgcgtt ggtgcaccag ggcgacgtgg tgcgtctccc caacatcacc	960
aacctgtgcc ccttcggcga ggtgttcaac gccaccaagt tccccgcgt gtacgcttgg	1020
gagcgcaaga agatcagcaa ctgcgtggcc gattacagcy tctgtacaa cagcaccttc	1080
ttcagcaact tcaagtcta cggcgtgagc gccaccagc tgaacgacct gtgcttcagc	1140
aacgtgtacg ccgacagctt cgtgttgaa ggcgacgacg tgcgcagat cgcctccggc	1200
cagacccggc tgatcgccga ttacaactac aagctgcccg atgacttcac gggctgcgtg	1260
ctggcctgga acacccgcga ctctgatgcc accagcacgc gcaactacaa ctacaaagac	1320
cgtaacctgc gccacgcgaa ctgcgcgcgc ttgcagcgcg atatcagcaa cgtgccttc	1380
agccccgatg gcaagccctg caaccccccc gccctgaact gttactggcc cctgaacgat	1440
tacggcttct acacacacac cggcatcgga tacacgacct accgctgggt ggtgctgagc	1500
ttcgagctgc tgaacgcgcc cgcacacgtg tgcggcccca agctgagcac cgaactgate	1560
aagaaccagt cgttgaactt caacttcaaa ggcctgacgc gccacggcgt gctgaccccc	1620
agcgcgaagc gcttcacgac cttccacagc ttggcgccgc acgtgagcga cttcacccag	1680
agcgtgcgcg atcccaagac cagcgagata ctgatatcaa gccctgcag cttcggcgcg	1740
gtgagcgtga tcaacccccg caccacgcgc agcagcgagg tggcctgtgt gtacacggac	1800
gtgaactgca ccgatgtgag ccccgccatc cagcgcgata agctgacccc cgcctggcgc	1860
atctacagca ccggcaacaa cgtgttcacg accacggcgc cgtgtctgat cggcgccgag	1920
catgttgaca ccagctacga gtgtgatata cccatcgccg ccggcctctg gccacgctac	1980
cataccgtga gcctgtgcgc cagcacccag cagaagacga tcttggctta caccatgagc	2040
ctggggcgcc	2049

<210> SEQ ID NO 69

<211> LENGTH: 1623

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<221> OTHER INFORMATION: Standardized optimization of TPA-S2 protein

<400> SEQUENCE: 69

atggtgcga tgaagcgcgc cctgtgtgt gtgctgtgc tgtgtggcgc cgtgttcgtg	60
agccccagcg ccgcgcgcag cggcgataga agcatgcct acagcaacaa caccatcgcc	120
atccccacca acttcagcat cagcatcacc accgaggtga tgcctgtgag catggccaa	180
accagcgttg attgaacat gtacatctgc ggcgacagca ccgagtgcc caaactgtgtg	240
ctgcagtacg gcagctcttg caaccagctg aacgcgcgcc tgaacggcat cgcgcgcgag	300
caggacgcga acacccgcga ggtgttcgcc cagtgaaagc agatgtacaa gaaccccacc	360
ctgaagtaet tggcgcgctt caacttcaag cagatcctgc ccgaacccct gaagcccacc	420
aagcgcagct tcatcgagga tctgtgttc aacaaggtga ccttggccga cgcgcgcttc	480
atgaagccag acggcgagtg cctggcgagc atcaacgccc gccacgtgat ctgcgcgcag	540
aagttcaag cgcctgacct gctgcgcccc ctgtgacgc atgaactgat cgcgcgcacc	600
accgcgcgcc tggtagcgcg caccgcgccc gcggcgtgga ccttcggcgc cggcgccgcc	660
ctgcagatcc ccttcgcgat gcagatggcc taacgcttca acggcctcgc cgtgacccag	720

-continued

aacgtgctgt acqgaacca gaagcagatc gccaacagt tcacaaagg catcagcaag	780
atccaggaga gactgaacc caccagacc gccctggga agctgcagga tctgggtaac	840
cagaagcccc aggcctgaa cccctgggt aagcagctga gcagcaact cgcgcaccac	900
agcagctgc tgaacgatat cctgagccgc ctgataagg tggagccga ggtgcagatc	960
gaacgcctga taccagccg cctgcagag ctgcagact acgtgaccca gcagctgac	1020
cgcgcgcgc agatccgcgc cagcgcaac ctgcgcgcca caagatgag cgagtgcgtg	1080
ctgggccaaga gcaagcgctt ggaattctgc ggcgaaggct accacctgat gagctcccc	1140
caggcgcccc cccagcgctt ggtgttctgc catgtgacct acgtgcccga ccaggagcgc	1200
aacttcacca ccgcgccgc catctgcac gagggcaagg cctacttccc ccgcgagggc	1260
gtgttcgtgt tcacagccac cagctggttc atcaccagc gcaactctt cagcccccag	1320
atcatcacca ccgacacac cttcgtgagc ggcactgcg acgtggtgat cgcgcatc	1380
aacacacccg tgtaagatcc cctgcagcc gagctggata gcttcagga ggaagtgac	1440
aagtaattca agaaccatac cagcccgat gtgagctgg gcgacatcag cgcgcatcac	1500
gcgcgctgg tgaacatcca gaaggagatc gatgcctga acgaggtgga caagaacctg	1560
aacgagagcc tgatgatctc gcaggagctg gccaagtac agcagtacat caagtggccc	1620
tgg	1623

1-434. (canceled)

435. An isolated polynucleotide comprising a nucleic acid fragment which encodes at least 20 contiguous amino acids of a SARS-CoV polypeptide selected from the group consisting of:

- (a) SEQ ID NO:2;
- (b) SEQ ID NO:4;
- (c) SEQ ID NO:6;
- (d) SEQ ID NO:8;
- (e) SEQ ID NO:10;
- (f) SEQ ID NO:12;
- (g) SEQ ID NO:14;
- (h) SEQ ID NO:16;
- (i) SEQ ID NO:17;
- (j) SEQ ID NO:19;
- (k) SEQ ID NO:21;
- (l) SEQ ID NO:23;
- (m) SEQ ID NO:56;
- (n) SEQ ID NO:58; or
- (o) SEQ ID NO:62;

wherein said nucleic acid fragment is a fragment of a human codon-optimized coding region encoding said SARS-CoV polypeptide, and wherein said human codon-optimized region is optimized by a method selected from the group consisting of: uniform optimi-

zation, full-optimization, minimal optimization or a combination of said methods.

436. The polynucleotide of claim 435, which encodes at least 50 contiguous amino acids.

437. The polynucleotide of claim 435, which encodes at least 100 contiguous amino acids.

438. The polynucleotide of claim 435, which encodes the complete SARS-CoV polypeptide selected from the group consisting of (a)-(o).

439. An isolated SARS-CoV polypeptide which is 90% identical to the polypeptide selected from the group consisting of:

- (a) SEQ ID NO:2;
- (b) SEQ ID NO:4;
- (c) SEQ ID NO:6;
- (d) SEQ ID NO:8;
- (e) SEQ ID NO:10;
- (f) SEQ ID NO:12;
- (g) SEQ ID NO:14;
- (h) SEQ ID NO:16;
- (i) SEQ ID NO:17;
- (j) SEQ ID NO:19;
- (k) SEQ ID NO:21;
- (l) SEQ ID NO:23;
- (m) SEQ ID NO:56;
- (n) SEQ ID NO:58; or
- (o) SEQ ID NO:62;

wherein said SARS-CoV polypeptide is produced from a nucleic acid comprising a human codon-optimized coding region, and wherein said human codon-optimized region is optimized by a method selected from the group consisting of: uniform optimization, full-optimization, minimal optimization or a combination of said methods.

440. The polypeptide of claim 439, wherein said polypeptide is 95% identical to the polypeptide selected from the group consisting of (a)-(o).

441. The polynucleotide of claim 435 further comprising a heterologous nucleic acid.

442. The polynucleotide of claim 441, wherein said heterologous nucleic acid encodes a heterologous polypeptide fused to said at least 20 contiguous amino acids encoded by said nucleic acid fragment.

443. The polynucleotide of claim 441, wherein said heterologous nucleic acid encodes at least 20 contiguous amino acids of a heterologous SARS-CoV polypeptide selected from the group consisting of (a)-(o).

444. The polynucleotide of claim 442, wherein said heterologous polypeptide comprises a small self assembly polypeptide, and wherein said heterologous polypeptide self assembles into multimers.

445. The polynucleotide of claim 442, wherein said heterologous polypeptide is a secretory signal peptide.

446. The polynucleotide of claim 435, which is DNA, and wherein said nucleic acid fragment is operably associated with a promoter.

447. The polynucleotide of claim 435, which is messenger RNA (mRNA).

448. A vector comprising the polynucleotide of claim 435.

449. The vector of claim 448, which is a plasmid.

450. A pharmaceutical composition comprising the polynucleotide of claim 435 and a carrier.

451. The pharmaceutical composition of claim 450, further comprising a component selected from the group consisting of an adjuvant and a transfection facilitating compound.

452. The composition of claim 451, wherein said adjuvant is selected from the group consisting of:

(α)-N-(3-aminopropyl)-N,N-dimethyl-2,3-bis(syn-9-tetradecenylethoxy)-1-propanaminium bromide (GAP-DMORIE) and a neutral lipid;

a cytokine;

mono-phosphoryl lipid A and trehalosidicorynomycolateAF (MPL-TDM);

a solubilized mono-phosphoryl lipid A formulation; and CRL1005/BAK.

453. The composition of claim 451, comprising the transfection facilitating compound (α)-N-(2-hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecenylethoxy)-1-propanaminium bromide (DMRIE).

454. The pharmaceutical composition of claim 450, further comprising a conventional vaccine component of SARS-CoV selected from the group consisting of inactivated virus, attenuated virus, a viral vector expressing an isolated SARS-CoV virus polypeptide, and an isolated polypeptide from a SARS-CoV virus protein, fragment, variant or derivative thereof and/or one or more polynucleotides comprising at least one coding region encoding a SARS-CoV polypeptide, or a fragment, variant, or derivative thereof.

455. A method for raising a detectable immune response to a SARS-CoV polypeptide, comprising administering to a vertebrate a polynucleotide of claim 435, wherein said polynucleotide is administered in an amount sufficient to elicit a detectable immune response to the encoded polypeptide.

456. A method for raising a detectable immune response to a SARS-CoV polypeptide, comprising administering to a vertebrate the composition of claim 450 in an amount sufficient to elicit a detectable immune response to the encoded polypeptide.

457. A method for raising a detectable immune response to a SARS-CoV polypeptide, comprising administering to a vertebrate the composition of claim 451 in an amount sufficient to elicit a detectable immune response to the encoded polypeptide.

458. A method for raising a detectable immune response to a SARS-CoV polypeptide, comprising administering to a vertebrate the composition of claim 454 in an amount sufficient to elicit a detectable immune response to the encoded polypeptide.

459. A method to treat or prevent SARS-CoV infection in a vertebrate comprising: administering to a vertebrate in need thereof the polynucleotide of claim 435.

460. A method to treat or prevent SARS-CoV infection in a vertebrate comprising: administering to a vertebrate in need thereof the pharmaceutical composition of 450.

461. A method to treat or prevent SARS-CoV infection in a vertebrate comprising: administering to a vertebrate in need thereof the pharmaceutical composition of 451.

462. A method to treat or prevent SARS-CoV infection in a vertebrate comprising: administering to a vertebrate in need thereof the pharmaceutical composition of 454.

463. A method of producing an isolated antibody, or fragment thereof, comprising administering the polynucleotide of claim 435 to a vertebrate and recovering said antibody or fragment thereof.

464. An isolated antibody produced by the method of claim 463.

* * * * *